PhD-FSTM-2021-095

The Faculty of Science, Technology and Medicine

DISSERTATION

Defence held on 09/12/2021 in Esch-sur-Alzette

to obtain the degree of

DOCTEUR DE L'UNIVERSITÉ DU LUXEMBOURG
EN BIOLOGIE

by

Silvia MARTINA

born on 31st of December 1990 in Copertino, (Italy)

CELLULAR HETEROGENEITY AS EMERGENT BEHAVIOR IN SYSTEMS BIOLOGY

Dissertation Defence Committee:

Prof. Dr. Alexander Skupin, Dissertation Supervisor (*Université du Luxembourg*)

Prof. Dr. Martin Falcke (*Max Delbrück Center for Molecular Medicine, Berlin, Germany*)

Prof. Dr. Massimiliano Esposito, Chairman (*Université du Luxembourg*)

Dr. Carsten Marr (*Helmholtz Center Munich, Munich, Germany*)

Prof. Dr. Etienne Fodor, Vice-Chairman (*Université du Luxembourg*)

UNIVERSITÉ DU
LUXEMBOURG

Luxembourg Centre
for Systems Biomedicine

Luxembourg National
Research Fund

Affidavit

I hereby confirm that the PhD thesis entitled "Cellular Heterogeneity as Emergent Behavior in Systems Biology" has been written independently and without any other sources than cited.

Luxembourg, Esch-sur-Alzette, October 2021

_____

Silvia Martina

Acknowledgements

I guess that nobody can say that the PhD path is straightforward, smooth and without complications. From my experience, I have through moments of discouragements, perplexity and confusion but I had special persons always from my side, whom I am infinitely grateful for different reasons. I have to thank my supervisor Prof. Alexander Skupin for his academic and human way to solve my doubts, clarify scientific aims and put me always in the best conditions to learn efficiently new concepts and to enforce my intuitions. My parents helped me in believing myself and overcoming difficulties as a challenge to improve myself. They have taught to me that failures don't qualify a person but just give the opportunity to grow. I have to thank my parents to be as I am. I have to thank my best friends: Elisa, Giorgio e Ugo for their passion in doing research. Their enthusiasm motivated me to be persistent and never give up. With them I shared most of my academic path and they will be ever the persons I admire most both scientifically and humanly. I have to thank my boyfriend, Daniele, for his love, presence and support anytime I needed. He gave me the serenity and the encouragement to work as my best convincing me about my worth as a person before as a scientist. I have to thank the lovely Susan, the best post-Doc colleagues Valerie and Thais and the sweet Francoise for valuable human and scientific discussions. They were always available to help me anyway. I have thank all ICS people to provide me a comfortable and friendly environment to enjoy science. Finally, I am thankful to my CET committee for the intense discussions and the hints helping me to understand how to achieve good science.

Abstract

A main signature of living systems is the high degree of spatiotemporal organization of their structures that results in the emergence of functions across the cell, tissue and organ levels. Cellular heterogeneity, which emerges from the effect of molecular variability to phenotype properties, is a main source of this complex and structural organization of life. Current approaches in developmental biology also suggest that cell-to-cell variability might be key to cell differentiation and investigate the underlying mechanisms. However, the impact of internal regulatory mechanisms of individual cells on the global population behavior and the role of the immanent stochasticity in cell-fate decision is still elusive.

This thesis applies complementary top-down and bottom-up systems biology approaches to investigate mechanisms of cell decisions and their impact on biomedical relevant systems. For this purpose, state-of-the-art single-cell data analysis approaches are combined with theoretical models and computational results to assess the role of intrinsic noise at the transcriptional level on cellular differentiation processes. The data driven top-down approaches revealed thereby previously unknown cellular heterogeneity in cell activation and brain composition potentially linked to diverse brain disorders. The complementary mechanistic-modeling and computational-based bottom-up approach provided new evidence for cell-fate decision as a population-driven mechanism resulting in the emergent behavior of multicellularity. This approach gives a new perspective on modeling cell fate commitment based on the interface between transcriptional variability and external induction signals.

The first part of this thesis is dedicated to the analysis of experimental single-cell RNA-sequencing (sc-RNAseq) data to quantify the effect of molecular fluctuations on the composition and dynamics of cell populations. In particular, sc-RNAseq data analysis allowed to dissect subpopulations in microglia activation and to identify disease associated molecular signatures. The importance of cell heterogeneity during brain development and its relation to the development of epilepsy was further investigated in zebrafish larvae by sc-RNAseq, which provided new insights into early mechanisms of epileptogenesis.

In the second part of this thesis, a mechanistic modeling-based approach was computationally developed

to study the mechanisms of the interplay between molecular stochasticity and nonlinear interactions of a genetic toggle-switch motif. The computational analyses of the model identified underlying mechanisms of the emergence of cellular heterogeneity from molecular noise to cell-cell interactions and ensembles of coupled stochastic entities.

Overall, the integrative approach of this thesis combining single-cell RNA sequencing data analysis approaches and mechanistic modelling gives a new prospective in understanding cell commitment by considering the interplay between transcriptional variability, external induction signals and population coupling between cells.

Table of Content

List of Figures

x

List of Abbreviation

Abbreviation	Full name
T.S.	Toggle Switch
ODEs	Ordinary Differential Equations
SDEs	Stochastic Differential Equations
FPE	Fokker-Planck Equation
DNA	Deoxyribonucleic Acid
PCA	Principal Component Analysis
DNA-seq	DNA Sequencing
Dropseq	Droplet-based Sequencing
pdf	Probability Density Function
mRNA	messenger ribonucleic acid
sc-RNAseq	Single Cell RNA-Sequencing
CV	Covariance
GRN	Gene Regulatory Network

Chapter 1

Introduction

Mankind has always been fascinated by life. The ways to address the essential reason for its existence in form of microorganisms, animals, plants, human body and how it has been sustained by inanimate matter like nutrients, sun and ultimately the universe, have been a focal point of life and natural sciences for centuries. Piece by piece, the scientific community is reconstructing this puzzle of life and this thesis intends to add a piece to our understanding of life. The main focus of this thesis is the cell differentiation process, which is essential to many physiological processes like the development of organs and organisms, cell reprogramming or the onset of cancer.

1.1 *Order-from-disorder* organization of life

Life is sustained by a multi-layer structural organization composed of different spatiotemporal scales of interactions connecting a plethora of entities that allows for fine-tuned regulatory processes. Living organism has to consist of a large number of atoms to bring out a highly ordered state as a collective phenomenon at large scale emerging from the randomness at small scales [1]. Each level of these local regulations can be pictured as an open system and a self-controlling complex network. This abstract vision has been recently investigated to study ecological systems [2], the organization of worldwide trade [3] and even in modelling social interaction [4] in various contexts.The multi-layer network structure is also a simplified scheme to understand the mechanisms underlying the biological organization of living matter [5]. This perspective is based on the fact that biological systems are regulated by an incredible number of entities and the interactions between them generate a network with high complexity. Specifically, biological complexity include networks of biochemical reactions and signalling pathways which compose the metabolic network,

the internal cell regulation by transcription and translation mechanisms, tissue and organism organization. Furthermore, the close cooperation of some components can give rise to emergent phenomena at a different length and/or timescale. These specialized spatial scale-depending levels communicate with each other to ensure a sustainable life for a multi-cellular organism.

One of the major challenges of modern biology is to bridge these scales ranging from the subcellular and cellular levels up to organs and tissues with the aim to discover hierarchical organization principles by identifying relationships between structure and function at multiple levels [6, 7]. The relatively new scientific approach of systems biology has the ambitious scope of explaining the basic mechanisms of molecular interplay and cell communication by revealing targeted functions, identifying patterns and finally elucidating the resulting multi-level complex structure from all those combined interactions.

1.2 Cellular heterogeneity and its role in differentiation

Cellular heterogeneity in complex living organisms is known to be one of the key element in biology to create functional cell populations for supporting the multiscale regulations and to maintain homeostatic states under different environmental conditions and against pertubations. A living cell, depending of its morphology, location and function in the body, is characterized by a specific phenotype whereas all cell types in our body share the same genetic code. It is really fascinating how nature assembles interacting cells to organize them into structures of higher order and patterns during embryogenesis and development by using a controlled spatial displacement [8, 9]. The basic mechanism which allows the multilayer arrangement of interactions is the occurrence of multiple cellular differentiation processes. Each involved cell responds dynamically to the external microenvironment and makes an adaptation decisions on its gene regulatory network [10]. This adaptive cell fate commitment seems to be triggered by a switch mechanism leading eventually to a desired cell state [11]. For many differentiation processes it is still unclear how the external signals and internal regulatory factors modulate each other to drive the switch between cell fates towards a specific cellular attractor [12]. Although some consequences of this targeted decision have been investigated like in an oriented cell division during animal development [13], in a tumor-growth environment [14] and in cellular reprogramming [15], a comprehensive understanding of the underlying mechanism is still lacking. The connection between different scales of interactions, the stochastic nature of intracellular biochemical reactions, the consequences for spatial cell organization and the generation of a wide-range of functional cells, makes an extensive study of cellular differentiation complex to investigate from both modeling and experimental points of view.

1.3 Cellular differentiation as an emergent behavior of cellular heterogeneity

Cell-cell variability represents a main mechanism for increasing complexity in a directional and flexible manner to assure on the on hand robustness an on the other hand adaptability in a cell population. The biological tendency towards tissue formation from a homogeneous population is therefore regulated by the graded differentiation of cells across spatiotemporal diversification and functional specification [16]. The underlying single-cell implementation (such as secretion of molecules [17], symmetric and asymmetric cell divisions [18], migration [4] and differentiation [19]) has an effect at the population level by generating an overall driving force of self-controlled organization which allows the emergence of a new complex architecture at the end. For the mechanistic understanding of this step, *symmetry breaking* is a key concept. In particular, the symmetry-breaking event occurs when, despite all cells being exposed to a uniform growth-promoting environment, only a fraction of cells becomes activated, differentiates and acquire new functions. The process is called 'symmetry breaking' because the transitions usually bring the system from a symmetric, but disordered and variable state, into one or more defined, less variable and asymmetric states (e.g. differentiated states) [20, 21].

This kind of collective behavior of several entities is investigated in diverse circumstances including ecological population [22], self-propelled colloidal motors [23, 24, 25] and in cancer cell populations [26] and is addressed by a large part of the scientific community due to its peculiarity in bridging different spatiotemporal scales. Mathematical models are been recently introduced to disclose the basic mechanism underlying this emergent phenomena. Statistical physics approach represent a powerful approach to conceptualize and assemble all these evidences into general first principle mechanisms to dissect potentially hidden heterogeneity dynamics [27]. Such concepts are particularly relevant in the context of cell differentiation. In this theoretical perspective cellular heterogeneity is the biological key able to create an adaptable environment where the interplay between dynamics and stochastic fluctuations becomes the essential ingredients to manifest phenotype variability.

Consequently, the combination of self-controlling feedback response as an individual decision and a collective behaviour resulting from cell-cell interactions might give a directionality towards a final stable state at the population level. Although cell fate choice during a differentiated lineage has been studied with different experimental [28, 29] and modelling approaches [30, 31], a broader perspective in which general concepts in statistical physics are applied is not yet completely investigated.

1.4 Context of this thesis

To validate and acquire a better understanding of the mechanism underlying cell differentiation, the thesis first investigates cellular heterogeneity of brain cells by single-cell RNA-sequencing analysis (Chapters 4 and 5) and subsequently introduces a theoretical approach to further investigate the stochastic component of phenotype variability in emergence of cellular heterogeneity at population level (Chapter 6).

The analyses of brain cell heterogeneity and its role in biomedical important systems focuses on (i) the diversity of microglia response up to stimulation, which is supposed to have an effect on neurodegenerative diseases Chapter 4, and (ii) brain development in zebrafish at single cell resolution for epileptogenesis study (Chapter 5). These analyses exploit the recent advancements in single-cell RNA-sequencing and applied state-of-the-art bioinformatic analysis tools to characterize cell fate dynamics in different conditions.

The starting point of the theoretical analysis in Chapter 6 is the widely studied regulatory motif of the genetic toggle-switch model. This mechanism describes the behavior of a flipped multistability system between stable states and reproduces the final choice of one of them [32]. A two-stage toggle switch analysis representing the transcriptional and translational levels of cells has been presented by Strasser et al. [33] in which the dynamics on discrete microstates has been investigated as an evolution of the state occupation probability in a Master equation approach. Since the biochemical reactions involve few molecules, the Gillespie's algorithm [34] has been used to simulate the stochastic behavior of the system based on its macroscopic mass action kinetics. The result was that the system evolves in the phase portrait according to a probabilistic flux throughout the intermediate metastable basins to reach a given attractor.

In this thesis, I investigate the stochastic nature of cell fate. For this purpose, I first applied a top-down approach by analysing single-cell RNA-sequencing data of the activation of microglia, which are the immanent immune cell of the brain, and changes in brain composition during epilepsy establishment in zebrafish. Both studies have revealed the stochastic nature of cell fate and identified key regulatory processes. Based on these findings, I used a bottom-up approach in the second part of the thesis to investigate mechanistically the effect of noise in a nonlinear dynamical system based on the toggle-switch model to dissect the interplay between a deterministic dynamics of two coupled variables (corresponding to two alternative phenotype specification) and noise. Performing computational simulations the probabilistic dynamics has been investigated and the system was characterized by its properties at different levels. This toggle-switch motif approach is not intended to explain the microscale molecular dynamics as an induced mechanism for cellular decision between two competitive final states but investigates the role of the stochasticity on the transcription level, the buffering effect of proteins on gene expression heterogeneity and the dynamics of coupled populations to have a more comprehensive view into cellular differentiation

and cell-fate decision.

Chapter 2

Aims and structure of the thesis

2.1 Motivation to use biology distributions for investigating cell differentiation

Molecular fluctuations operate in a mysterious and complex way. In the cellular differentiation process they might be the key for its comprehensive underlying mechanisms. At single-cell level the noise that acts predominantly on the transcription level might be minimized in order to strengthen specific molecular profiles and lead to the phenotype manifestations of a cell state. However, the same noise might also be part of the mechanism involved in the decision-making process in a proactive way [35]. These apparently conflicting roles displayed by this noise might be resolved by considering the underpinning gene regulatory network dynamics. The gene regulatory network is supposed to be subject to a rapid rearrangement when it is perturbed enough to be pushed far from an unstable state. In other words, a core of a subnetwork of genes might be activated by reinforcing specific pathways when the fate-decision occurs [36, 37, 38]. This state might correspond to an initial destabilization of the progenitor state. The cell is deemed to explore all possible gene configurations, each of them potentially associated to just one specific lineage attractor. When an external differentiation signal gives an instruction towards a particular phenotype, the cell reacts intrinsically according to its temporary and sensitive microstate and finally selects one of the predestined states. This supposed synergy between *instruction* (gene programs of differentiation) and *selection* (individual cell choice) might be a mechanistic explanation of the experimental detection of "rebellious cells" in the commitment of blood progenitors cells to erythroid or the myeloid lineage [39]. These cells follow the fate opposite to the one intended by the differentiation signal due to gene expression fluctuations. Such fluctuations could unexpectedly overcome the bias signal and allow cells to enter the

"non-intended" attractor basin. In the theoretical part of this thesis, I based my bottom-up approach of modelling cellular differentiation process on the biology distributions framework. The models which I introduce in Chapter 6 are addressed to give insights into the underlying mechanisms, involved during cellular differentiation, by matching the dynamics into the supposed preexisting gene program and the stochasticity into the gene expression variability. My idea of how the cellular differentiation is accomplished in terms of biology distributions evolution is shown in Fig. 2.1.
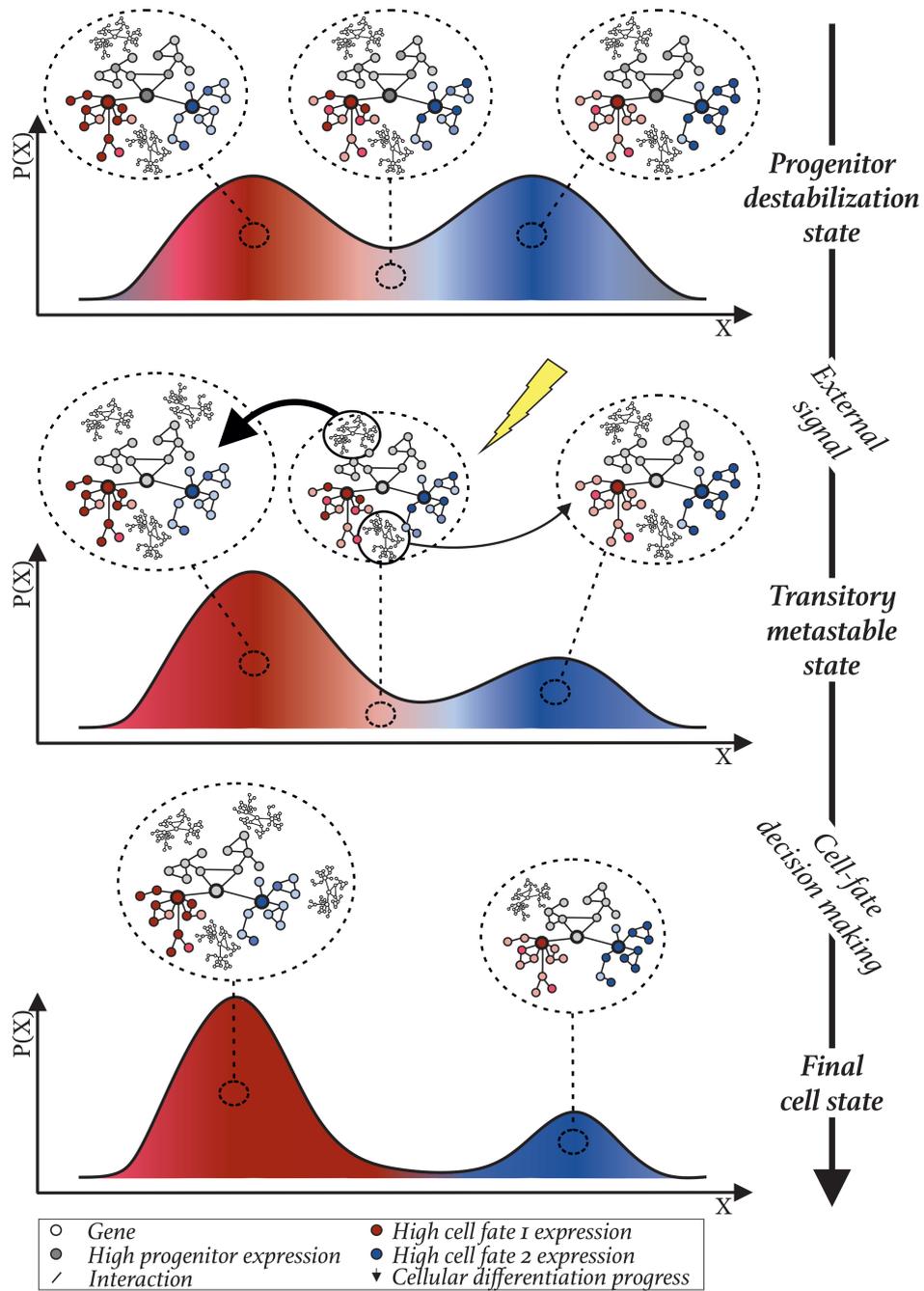
Figure 2.1: *The preliminary stage associated to the progenitor destabilization state is characterized by a symmetric probability distribution of two allowable cell states. This macroscopic description emerges at the population level from the equal probabilistic repartition in two main sets of gene regulatory networks of all possible internal cell microstates. This ambiguous state is prone to be rapidly rearranged by external signals. This transitory state, triggered towards a predominant final cell state, is macroscopically dependent on the all individual cell-fate decisions, which eventually define the final cell state probability distribution at larger scale.*

2.2 Specific aims of the thesis

This cumulative thesis addresses cellular heterogeneity in the context of differentiation by an interdisciplinary approach and (i) performs state-of-the-art data analysis of single-cell RNA-sequencing data to reveal cellular heterogeneity of brain cells and (ii) introduces a new generic model for cell fate induction to investigate underlying mechanisms.

In particular, the thesis focuses on

1. Microglia diversity upon stimulation (Chapter 4).
   Molecular variability was untangled in the study of the microglia phenotype diversity under an experimental treatment, which induced microglia activation compared to a saline condition.

2. Cellular diversity in brain development of zebrafish and its impact during the establishment of epileptic seizure (Chapter 5).
   Biological heterogeneity dynamics during brain development was investigated in an epileptic zebrafish model with a focus on molecular signatures of epileptogenesis .

3. Mechanistic modelling of cell differentiation with a focus on noise driven cell fate (Chapter 6).
   Based on the sc-RNAseq data analysis approaches in Chapters 4 and 5, a generic model for cell fate induction was established and comprehensively analyzed.

2.3 Structure of the thesis

Chapter 3 provides a general overview of the main concepts and tools applied in this thesis. The subsequent two chapters present the published papers in collaboration with experimental groups addressing molecular variability and cellular heterogeneity with direct clinical implications. My contribution to these studies was mainly based on the analysis of single-cell RNA sequencing data, which was able to identify underlying mechanisms of cell differentiation in health and disease conditions. Single-cell RNA-sequencing data analysis was applied as a first approach to the study of emergent phenomenon of cellular heterogeneity. It allowed me to understand the several sources of molecular variability and how to infer reliable information by pooling data from different conditions and time-points in microglia activation (Chapter 4 and brain development during epilepsy establishment (Chapter 5). Remarkably, these two studies have highlighted the relevance of single-cell experiments to be a valid support for acquiring more insights at the molecular level and broaden the knowledge of intrinsic underlying mechanisms. Chapter 6 is entirely dedicated to my extensive bottom-up study in developing suitable models for revealing mechanistic properties, that might

characterize the cellular differentiation process at its main stages. In Chapter 7, I recapitulate the key points of the thesis and how this work is contextualized in the current challenges in System Biology.

Chapter 3

Scientific Background

This chapter gives the scientific background of the thesis by introducing the main concepts of cellular heterogeneity and experimental and theoretical approaches applied in an interdisciplinary manner in the following chapters.

3.1 Bio-complexity in a multi-cellular organism

There are about 200 different type [40] of cells in the human body like neurons, red blood cells, skin cells, astrocytes, immune cells and many more. The cell functions associated to them are diverse and include structural support for tissues, immune response against pathogens and energetic support for metabolic processes. The specification to carry out a specific function is characterized by a particular cell morphology and molecular orchestration which typically depend on the location in the body. The cell variability is organized by different levels of regulations and interactions, which are interconnected and modulated by each other to sustain the survival of the organism as a whole.

The active maintenance of all biological tasks needed for survival relies on a complex stratified network structure. The links between all the components are varying dynamically and have to be locally flexible to allow for slight and sudden changes without compromising the system's sustainability. The various transporters and channels (voltage-gated, ionic, ligand-gated, mechanically-gated) and the corresponding timing of these connections are key elements for communicating messages efficiently between specific compartments, cells, organs and cellular systems. Therefore, a functional organism requires a huge number of different entities operating in a orchestrated and cooperative way to rapidly exchange information by using diverse communication mechanisms. An alteration of this dynamical equilibrium could dramatically

lead to the onset of cell-growth related diseases such as cancer or to cell-type specific cell death as e.g. in neurodegeneration. Although the initial defect can involve only one cell, the biological environment and other structural modifications could spread the error over the entire local cell-type population and potentially in other parts of the body. Thereby the physiological conditions, the spatial local structure and the proficiency of cells to communicate with each other are the main factors affecting the self-sustained micro-environment. In addition, biological complexity is also induced by the huge number of entities assuring a persistent living state for a multicellular organism. For instance a human body is composed of roughly 3×10^{13} cells and each of them is an open system which continuously cooperates and interacts with its neighbors to maintain homeostatic condition under several conditions and perturbations.

3.2   System biology

To disentangle the biological complexity, systems biology was established as a new interdisciplinary approach in the beginning of the millennium [41]. One of the current challenges in system biology is to tackle the plethora of entities interacting over different spatiotemporal scales and to combine these efficiently into a solid multi-scale framework from large collections of multi-omics and other data for mechanistic insights into biomedical systems.
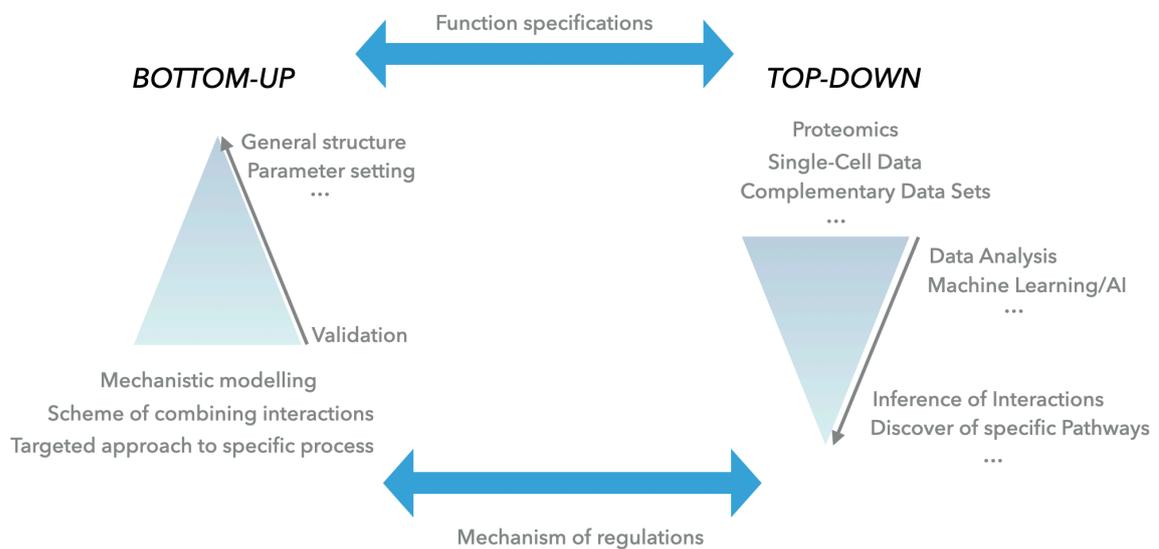
Figure 3.1: *Schematic representation of the methods used in system biology to bridge biological scales and the way to integrate them in a coherent framework.*

To disentangle the interplay of the different spatiotemporal biological, systems biology follows to comple-

13

mentary approaches in order to navigate between the different levels of descriptions, namely the bottom-up and the top-down approaches as represented schematically in Fig. 3.1.

Intercellular interactions, which are at the heart of this thesis, can be used to exemplify these concepts. Cell-cell communication can occur via many different intercellular signalling pathways depending on the cell type or cell state, and the detailed chemical and/or physical mechanisms of a given pathway, which are investigated by theoretical and experimental approaches [42]. Following a bottom-up approach, the individual mechanisms can be combined in a model to examine the global behavior of a cell or a cell population resulting from the different interactions. Alternatively, top-down approaches are based on observed hallmarks of cell populations in order to gain insight into their working principles. In modeling-based top-down approaches, very simplified mathematical models - also referred to as toy models - are used to mimic the dynamics of the population and do not aim at reproducing specific molecular mechanisms. These models are addressed to capture universal laws by linking phenomena of a different nature.

Therefore, many strategies are adopted to tackle biological complexity. To get reliable understanding of a biological system, combined approaches have to be carefully selected according to the posed question. The spatiotemporal scales of interest thus have to be chosen adequately, depending on the required level of accuracy and the availability of experimental and/or theoretical tools. Complementary tools such as statistical analyses [43] and machine learning techniques that allow to infer information from Omics big data [44] are also commonly used by the theoretical scientific community to untangle the complexity of biological systems. Depending on the complexity of the analysis describing a biological system, a complete analytic description might need a large-scale integration over multi-layers omics data. This methodology allows to cross-validate the results from several techniques and study the faithfulness of a mathematical model in reflecting partially or globally the experimental evidence as well as its predictive power under different conditions. On the other hand, it could be tricky to extrapolate reliable information and interpret it correctly due to the huge amount of potentially inter-correlated data. The high-throughput data has to be polished from inconsistencies and technical readouts before making inferences by the integration over different scale levels.

3.3 Single-cell analysis and single-cell RNA-sequencing technology

Single-cell analysis has already widely demonstrated to be a powerful top-down approach to investigate cellular dynamics and therefore discern populations of cells from their heterogeneity in function and fate [45]. The gathered molecular knowledge of a biological process can then provide mechanistic insights into individual cells and their interactions can be also inferred from the integration analysis over multi-omics

data providing a more comprehensive understanding of the involved mechanisms [46, 47]. RNA-sequencing (RNA-seq) is one of a recently developed approach that allows an unbiased inspection of the transcriptome [48] [49] [50]. The phenotype information of each cell allows to identify previously uncharacterized molecular heterogeneity at the cellular level and dissecting gene-function relationships. Basically, the monitoring of the transcriptional activity in single cells by measuring directly messenger RNA levels, allows to identify each cell in the high-dimensional gene space.

A major challenge in cell-to-cell variability analysis from sc-RNAseq data is the mix between the biological stochastic transcriptional activity and the measurement noise. Indeed single-cell gene expression data contain an abundance of dropouts that leads to zero expression measurements as well as other technical errors that may confound distinctive transcription factor expression states [51, 52]. Advanced methods of statistical inference are needed to first identify a reduced gene set (by basically exploiting co-expression patterns) in order to lower the space dimensionality and facilitate data visualization. Secondly, the inference is required to interpret appropriately the variability of gene expression counts and produce reliable outcomes [53]. This stochastic nature can be partially explained by noting that gene expression levels are regulated by combinatorial interactions between numerous cellular components, where these interactions involve random biochemical reactions [54, 55, 56].

Before the recently established methods for spatial transcriptomics [57], the limited amount of spatial information in typical single-cell RNA-sequencing experiments is one reason why individual methods have to be integrated to quantify the processes occurring in the cell from signal recognition to cell fate [45]. The dynamic processes in living cells are schematically summarized in Fig. 3.2 (David G.Spiller et al. 2010).
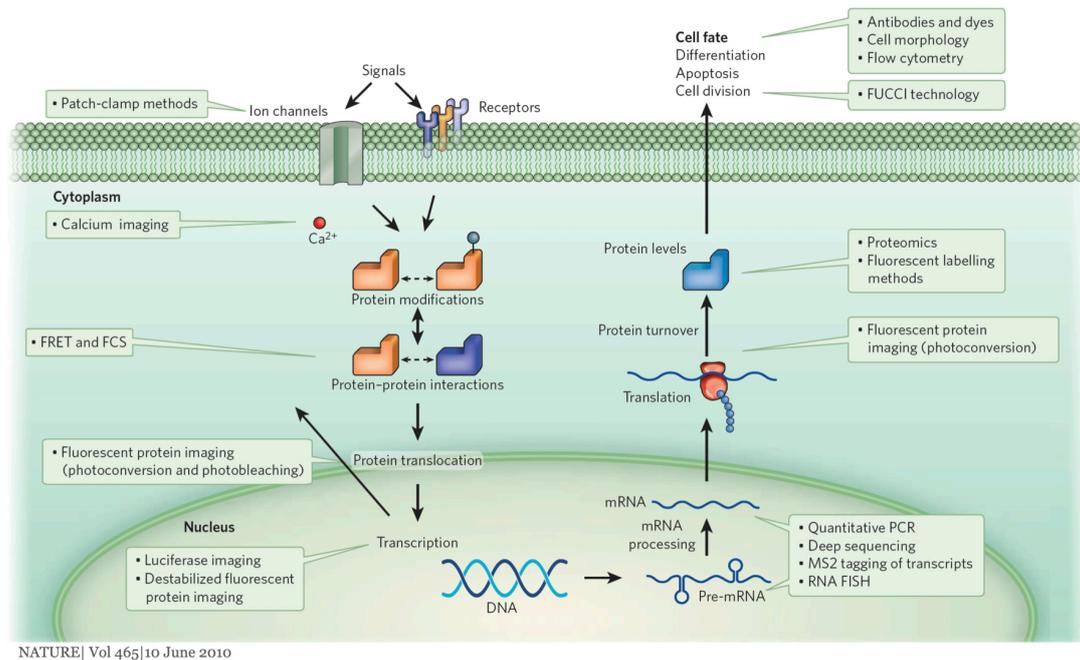
Figure 3.2: *Different experimental tools are required to collect measurements from different stages in the regulation of cell signaling, transcription, plasticity and fate in single cells (picture from [45]).*

Even if all the relevant molecular measurements could be made, the complexity of the data makes it difficult to approach systematically, integrate coherently and interpret correctly the hidden information. For instance, cellular heterogeneity is a feature that is intrinsic to many cell-fate processes, including division, apoptosis [35] and the generation of induced pluripotent stem (iPS) cells [58]. Furthermore, intrinsic stochastic events which are caused by the small number of molecules and genes copies make cell signaling and transcription governed by noise. In order to study the intrinsic heterogeneity of a process, the involved cells have to be identified by their common time-depending gene level expressions like for example in studies of the cell cycle [59], the cardiac clock [60], transcriptional cycles [61] and cell-signalling dynamics [62]. Often, cell synchronization is neither possible to achieve nor a desired condition when studying physiologically relevant processes. Indeed, the main mechanism of cell-to-cell variability emergence could be inferred from the asynchronous dynamics of cells.

The inherent presence of random noise in single-cell RNA-seq data account also for different hidden factors that might result in gene expression heterogeneity, and an appropriate computational analysis could reveal hidden subpopulations of cells [63]. Characterizing quantitatively transcriptional dynamics would give additional information of the change of gene expression programs over time, differentiation and development and in response to stimuli [64]. Therefore, by studying temporal transcriptional regulation programs,

transient cell states might be identified by the asynchronous cellular responses to either developmental cues or environmental perturbations [65]. Transcriptional levels are a powerful indicator of the state of an individual cell at a point in time because it is defined by the instantaneous gene-gene correlations. The collection of few static snapshots of the dynamics renders the inference of causality difficult to achieve. The recently established approach of *RNA velocity* addresses this challenge by providing additional information to support the prediction of the future state of individual cells on a timescale of hours by introducing a differential equation for the spliced mRNAs in relation with the unspliced molecules [66]. Hence, the high-throughput single-cell RNA-seq data is a valid source for dynamically investigating molecular processes that underlie cell-fate decision in single-cells, understanding cell plasticity under different stimulus and elucidating developmental pathways [67].

3.4 The gene regulatory network and its relation to cellular differentiation

The mRNA expression levels are indicative of the corresponding gene expressions and therefore can describe the cell state. Gene regulatory network are logical models that partially reconstruct the most relevant interactions between genes deduced from specific transcription factors and complementary experimental data by computational modeling approaches [68]. In this context, single-cell transcriptomic data has provided a huge basin of applications into reconstructing gene regulatory networks [69, 70, 71]. Since underpinning the cellular phenotype, changes in the gene regulatory network describe the temporal evolution of the cell state where changes in gene interactions over time occur due to internal regulatory mechanisms and external stimulus. Hence, the heterogeneity could be dissected from data of different conditions and time points based on the gene regulatory network. This analysis allows for a better understanding of the basic variation that characterizes dysregulated cellular process and shedding new light on disease mechanisms [72].

Computational methods have been developed to statistically characterize gene expression distribution for understanding expression variation and for selecting marker genes for cell population identities [73, 74]. Statistical analysis based on cell similarity have contributed to identify cell types plugging into the framework of a hierarchical clustering [75] as well as information theory based methods to detect genetic interactions [76] and core regulatory circuits [77]. The dynamical reconstruction of the topology of interactions in high-dimensional data has been also widely investigated [78] in order to uncover the underlying non-linear dynamics and making accurate predictions based on time-series datasets [79].

One of the major interests in the study of the evolution of gene regulatory networks is in inferring direct pathways along cellular differentiation based on identifying stage-dependent cellular commitments by causal connections between transcription factors. In this framework, hubs in gene regulatory network provide information about the internal driving-force that moves the entire network into a stable configuration underlying a specific phenotype as a cell state. During this process, the gene regulatory network evolves not by following only the signal transduction pathways of differentiation but also according to many unpredictable and untraceable factors like the burst cellular reactions to sudden external stimulus [80, 81], the mechanic random displacement of molecules and the opening of unexpected new pathways [82]. All these contributions result in a heterogeneous cellular responses due to the individual activated mechanisms and dynamical molecular behavior [83]. Consequently, the cascade of multiple changes in gene interaction and the resulting cell state evolution is then hardly to be revealed. Despite these obvious limitations, single-cell transcriptomics have provided detailed phenotypic information to predict lineage trajectories [84], to perform a correlation analysis within and between datasets to discriminate lineage-specific genes [83] and to reveal dynamic regulatory relationships finally conducting to the reconstruction of direct protein-protein interaction network [85].

3.5   Modeling cellular heterogeneity

It has been widely demonstrated that cell-cell variability is deeply embedded in the inherently stochasticity of the gene expression process [86, 87, 55, 88, 89]. The intrinsic molecular and cellular sources for such variability [90] have been associated to either biophysical processes (e.g. the random partitioning during mitosis [91]) or biochemical regulation (e.g. the dynamical functioning of the intracellular network [92] or the chromatic dynamics [93]). In order to infer biological noise from complex high-dimensional data sets, mathematical modeling and model-based data analysis are required to process, integrate and interpret them properly. Models are fundamentally important to deduce the design principles characterizing complex nonlinear behavior in space and time such that patterns, oscillations, switching and stochasticity. In particular, since single-cell data are fundamentally noisy,stochastic methods need to be used in combination with deterministic models to respect cell-to-cell variation and the immanent noise [45] in mechanistic approaches to cellular heterogeneity [62]. The intrinsic noise that occurs in transcription [86] and translation [94] is often incorporated in stochastic simulations by the Gillespie algorithm and its derivatives [95]. In this approach, a simplified mechanistic description assumes individual molecular reactions occurs following an exponentially distributed half-lives (waiting-times). Using stochastic differential equations (SDE) is also a valid approach to investigate noise effect in case a minimum number of molecules is present and then

applicable in real cellular systems. Data sets of time course of single-cell snapshot data have offered the possibility to study stochastic transcriptional dynamics and validate stochastic models. Temporal dynamical connections between discrete and static information might be a valid tool for inferring causality between transcription factors, elucidating pathways for cell lineage specification and capturing the variation over progression of cell-to-cell variability.

The final integration of experimental methodologies and modeling approaches across different spatiotemporal levels is eventually the only conclusive way to elucidate biological processes by looking at it from a comprehensive perspective. Combining information from several fields and trying to make coherent and predictive conclusions is the ultimate challenge in biology systems [96, 97].

3.6 Models for cellular differentiation

Cellular differentiation is the fundamental biological process which drives the embryonic development from the egg as single-cell to the multicellular organism by progressively allowing the creation of cells tailored towards specific functions [98]. Hence, this process is the biological mean to increase phenotype variability in order to produce and place specific cell-types.

The main features of cell differentiation have been identified in *stability*, *directionality*, *branching*, *exclusivity* and *promiscuous expression* as Foster et al. (2009) have claimed to build their model of sequential branching in hierarchical cell fate determination [99]. Stability is referred to the presence of multiple stable states in a system dynamics description. Indeed, a specific cell type is thought to be established when a stationary stable state is reached, which is also referred to as cellular attractor. Directionality is based on experimental assessments of the progressive pluripotency reduction guiding the progression from progenitor cells to more specialized cell types [100] that only rarely regain pluripotency by de-differentiating [101]. Subsequent binary cell-fate decisions ramify in alternative pathways on the developmental structure. Each committed stage to a specific lineage excludes the possibility to access the other branch in the subsequent stages. The multipotent progenitor cells co-express specific genes related to all their associated lineages. The promiscuous gene expressions, before a strong core gene activation is accomplished, might be a intermediate metastable state sensitive to perturbations. This minimal proposed model contains all the key ingredients that seem to govern the cell phenotype development.

The most common systematic procedure to investigate the mechanism which control cell fate decisions is to build a model with multistability in the dynamics based on ordinary differentiation equation describing interaction between specific transcription factors in agreement with experimental observations [100, 102]. Therefore the dynamics is investigated according to the parameter choice that modulates the size of the

region where monostability, bistability and in general multistability are localized. Since the choice of the steady state is not only regulated by the initial condition but also by signalling, the model should take into consideration secondary pathways controlled for instance by cell polarity, cell adhesion and cell division that leads to a spatially organized arrangement of the cells. Indeed, stochastic spatial simulation indicate that uneven repartitions of molecules at division are also able to break the symmetry and to prime specification [96]. The noisy molecular fluctuations can be investigated by its contribution to affect robustness in a multi-stable regime. A model taking into account these intercellular signaling and cross-signaling between the cells has been developed by Joëlle De Caluwé et al. in [103]. Recently a computational model based on cell intrinsic kinetics has revealed other possible hematopoietic lineage pathways even without considering cell culture environment in which other factors like competition in space or growth factor might provide a more complete description [104].

In a more mechanistic-based modeling approach for cellular differentiation, different models focusing on the stochastic gene expression have been proposed. In a first model, stochasticity in gene expression is caused by random displacement of regulators along the DNA. Phosphorylation and/or dephosphorylation of transcriptional regulators are triggered by signal transduction between cells and responsible for the stabilization of gene expression [98]. In another model, fluctuations in regulatory signals are also considered for the increase of the population variability which is modulated by negative and positive feedback loops [55]. A third model encourages the use of conceptual tools to explain emergent cell behavior from gene expression. Indeed, a more extended view which jointly embraces the concepts of state space, high-dimensionality and cellular heterogeneity could give a new view of explanations based on principles of statistical physics in a mathematical framework [105].

3.7 Waddington landscape as an abstract framework to understand cellular differentiation

Conrad Waddington introduced the notion of the 'epigenetic landscape' as an abstract methaphor to visualize and conceptualize the emergence of developmental choices [106, 107, 108].
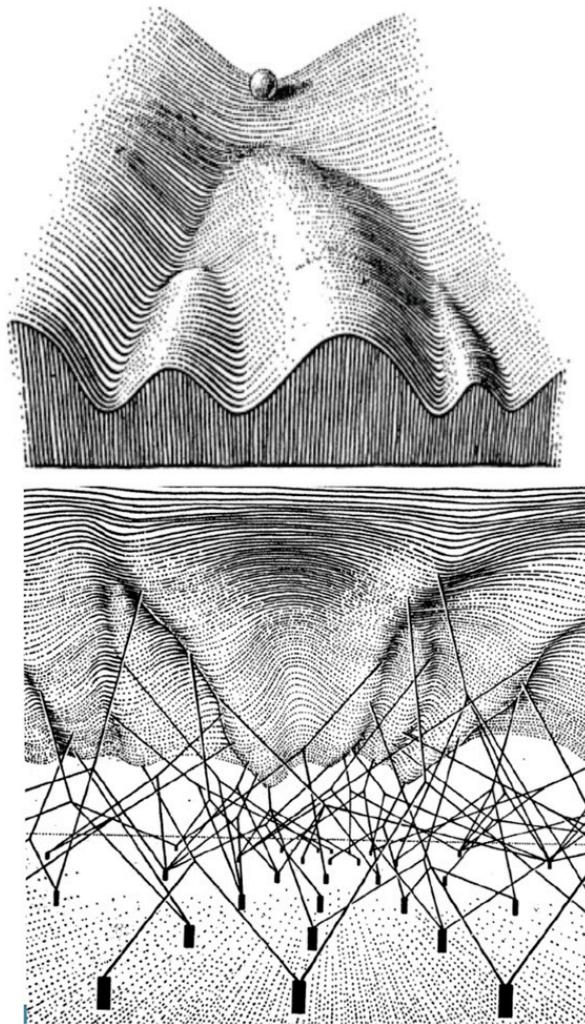
Figure 3.3: *Waddington epigenetic landscape in its classical representation introduced by Conrad Waddington (picture from [108]). A cell is represented by a ball, starts differentiation at the top of a hill and rolls down the landscape through a series of branching points, that represent the cell decision event. Underpinning the landscape the gene activity constructs the valleys.*

In this vision, the landscape consists of a hill with a series of branching points where the cell as a pebble rolls down and, according to the sequential binary directions taken, finally reaches a specific cell state as a cellular attractor. In a modern interpretation of this representation, the gene activity is pictured as a peg support underpinning the valleys and the hills which result in the *quasi-potential landscape* [109, 110]. The quasi-potential landscape is in principle built as an extra dimension on the multi-dimensional phase space of the gene expression by evaluating at each point the vector field that governs the direction of the instantaneous change of the state. Specifically the potential is the scalar quantity obtained by the physics

relationship between forces and potential:

$$\overrightarrow{F}(x_1, x_2, ..., x_N) = -\nabla U(x_1, x_2, ..., x_N).$$ (3.1)

In the multi-dimensional system the vector field is often not the gradient of some function [111] and genetic networks belong to these class of non-equilibrium systems in which the energy is not conserved necessarily [110]. The attempt to construct a landscape in the approximation of quasi-steady state assumption has given it the adjective of a 'quasi-potential'. This theoretical approximation is motivated by the possibility to move the theoretical description into a more appealing framework of statistical mechanics where a formulation in term of equilibrium probability distributions of the system states can be introduced. However, living systems continuously operate in out-of-equilibrium conditions and a theoretical extension in the thermodynamics notions of energy, work and dissipation is required to define new reliable quantities in the description of non-equilibrium phenomena [112, 113].

Assuming that such potential is established, there are two means by which the cell might change its state [102]. The first way is determined by stochastic fluctuations that allow cells for discrete random displacements around the landscape and occasionally might show significant jumps into another attractor basin, causing a spontaneous state-switching event. Biologically, this could happen due to a more effective stochastic boost which promotes cell identity changes by overcoming a threshold level [114]. The landscape geometry affects the probability according to the deep attractor basins or the barrier length between basins making the probability of transition states dependent on the constrains defined by the underlying genetic network and its directional regulatory pathways [110, 115].

Alternatively, the induction of a state change can be caused by structural modification of the landscape due to the variation of some parameters of the system's dynamics. They are associated to the extracellular signalling that produce sudden qualitative change in the systematic behavior [116]. These abrupt parameter-depending changes in the stability regime are related to the bifurcation phenomena in dynamical systems. In both explanatory scenarios, the underlying assumption is that the cells are prone to follow a genetic programs formalized by a multidimensional dynamics established at the beginning of the cellular differentiation stage. The individual cell fate decision is therefore resolved either as a discontinuous, stochastic transition event modulated by signals [117] or a dynamical "jump" in the cell activities between discrete cell fates in a probabilistic matter[118].

Although this framework currently provides a rather general vision of cellular differentiation, the quasi-steady state assumption of the epigenetic landscape and corresponding static potential simplifies the underlying self-organization of the biological processes. In this perspective, the intra-population communication

of cells is discarded and consequently, one might loose the opportunity to study the cell-fate decision as an individual random event able to generate a global response at the population level. To address this limitations, I will focus in Chapter 6 on a slightly different perspective in which the quasi-potential landscape might change dramatically over time to respect the fact that cells operate in a non-equilibrium regime and actively shape the microenvironment determining the epigenetic landscape. Supposing a pre-existing stochastic gene expression dynamics, at a certain point the internal configuration might be driven by external signals, and a heterogeneous response for cell commitment might arise. Specifically, I believe that the phase which anticipates the cell decision is dominated by the stochastic nature of gene expression variability which allows random navigation through the phase space determined by the admissible gene networks that are equally likely at this stage. Subsequently, if an external signal creates a sufficiently strong stimulus in this probabilistic spatial investigation, it might lead towards a final stationary probability distribution for the phenotype expression. The corresponding shape would thereby depend on each individual single cell selection of the predestined states [119, 120, 121, 122, 123].

3.8 Cellular heterogeneity description by distribution biology

A fundamental function of heterogeneity is to provide robustness to a biological system when responding promptly to environmental perturbations in order to maintain a homeostatic equilibrium. Referring to microorganisms as a specific example, the spread of phenotypic traits within a population offers more chances of survival as a whole organism and adapt to different conditions including competition for resources. Although this adaptability might lead to the progressive disappearance of some traits due to their contradicting co-existence and/or the inadequacy in being an advantage in this new stable configuration of the population, some other mutations during this evolutionary process contribute to create a more sustainable environment guaranteed by the diversity of different organisms. This central mechanism of life has led to the general perception that *"Nothing in Biology Makes Sense Except in the Light of Evolution"* as stated by Theodosius Dobzhansky in his landmark assay in 1973 [124]. This conclusion can be generalized at the level of cell populations by making a correspondence between the biodiversity within ecosystem communities and the multicellularity in a living multicellular organism. The mutation and the interplay of induction and selection, which are the two major components underlying the life evolution mechanism, have now become into the heterogeneity and intra- and extra-cellular signaling concepts, respectively. By this association, the way to answer the fundamental question on how the multi-cellular organization emerges from the same genome in a coordinated matter [125, 126] in one individual is fundamentally linked to the life mechanism to increase complexity in a multi-particle interacting system. By balancing resilience [127]

and flexibility [128] the final aim of biology is to increase survival probability in a multicellular organism. Hence, understanding life across its different scales relies on investigating the *cellular nanoevolution*–that is the dynamics of cellular heterogeneity [27]. The origin of cellular heterogeneity (even within clonal populations) is originating through the multiscale organization of life as depicted in Fig. 3.4 taken from [27].
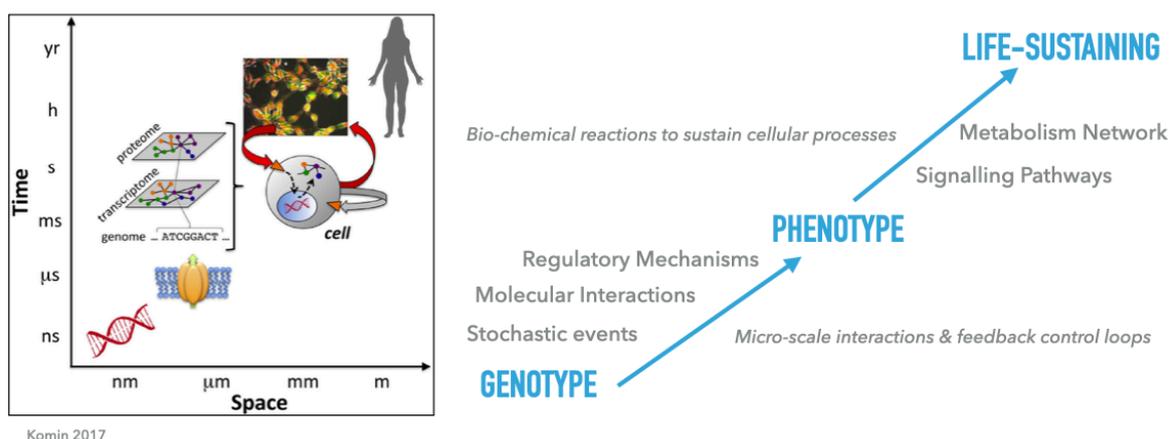
Figure 3.4: *Life-sustaining organization in multiple interactions at the same scale and across multi-layers communication. The biological information crosses over different spatiotemporal scales in a coordinated, synchronous and targeted matter to generate tissues, organs and organisms (picture from [27]).*

On the smallest relevant scales, the stochastic nature of molecular interactions induces individual transcription profiles that are subsequently instructed and selected on the level of the cell populations by intercellular signaling or cell-cell interactions leading to a coordinated generation of tissues, organs and organisms [27]. This underlying noise-driven cellular heterogeneity is the mechanism to balance robustness and adaptability [129].

In order to characterize cellular heterogeneity, I need to investigate biology at single-cell resolution. The cell is a micro-information progressing entity that has to respond to extracellular signals and consequently changes its internal state. A non-equilibrium condition is therefore a life-sustaining requirement, for instance, in enabling active signal mechanisms by ion gradient dynamics. The complexity of a biological system is encompassed by different nature of fluctuations that ranges from genetic heterogeneity [130], noise induced cell-to-cell variability on the transcription and translation level [39] as well as the individual behavior that may originate from specific subcellular organization [131]. To disentangle cellular heterogeneity over different sources of variability and uncover the underlying mechanisms that underpin biological complexity, it is necessary to rely on experimental cell-targeted approaches and subsequent integrative frameworks. Fortunately, the recent experimental technology in single-cell analysis allows to generate high-

throughput data enabling the development of a *distribution biology* framework following non-equilibrium statistical physics methods [129].

Chapter 4

Microglia diversity

This chapter includes the manuscript "*Single-cell transcriptomics reveals distinct inflamation-induced microglia signatures*" published in collaboration with the Michelucci lab from the Luxembourg Institute of Health in EMBO Reports [132] to which I contributed by advanced single cell RNA-seq analyses. The overarching aim of this study was to investigate the diversity of the activation response of microglia, which as the immanent immune cells of the brain contribute to diverse brain disorders including neurodegeneration.

4.1 Microglia activation as cell-type specific example of phenotype diversity

In the central nervous system (CNS) there are two broad classes of cells: neurons, which process diverse phsiological information, and glia cells, which provide the neurons with mechanical and metabolic support [133]. Microglia belong to the second type and are specialized in intervening promptly to the defence of the neural environment. Dysfunctional behavior of microglia are thought to aggravate CNS diseases such as Alzheimer's [134] disease and Parkinson's disease [135].

Microglia change rapidly and widely their gene expressions to modify their functional profiles accordingly to the received signals from their CNS environment. The inflammatory response of microglia is worth to be investigated deeply as they provide a primary, transient and self-limiting defence intervention by which tissue damage is repaired and harmful stimuli are resolved . Moreover, alteration of this mechanism leads to pathophysiological features present commonly in several neurodegenerative diseases. Identifying molecular profiles associated to specific functional activities in response to different CNS perturbations might give rise to more targeted therapeutic approaches to restore abnormal or detrimental microglial phenotypes found

in several CNS diseases.

In this work, LPS injection was used in a mouse model to mimic infectious condition and induce an immune system response to elucidate microglia heterogeneity towards systematic inflammation. Our study was based on a combination of single-cell RNA sequencing and multicolour flow cytometry techniques. My specific contribution in single-cell data analysis has allowed to characterize distinct activated states from their molecular profiles under acute inflammatory conditions. After having exploited qPCR approach to select cells under steady state (saline condition) and LPS treatment, single-cell experiments were performed by using the recently developed high-throughput droplet-based Drop-seq method [136]. The resulting amount of data to be processed was 1247 cells (477 from the saline and 770 from LPS-injected mice) with 12,369 expressed genes.

For the comprehensive analysis, I followed the state-of.the-art analysis workflow for sc-RNAseq data [137]. First, PCA was used for dimensionality reduction and clustering. Subsequently, I performed differential expressed gene (DEG) analysis to confirm the clustering procedure by identifying cell states based on the expected transcriptional profiles for the up- and down-regulated genes in cells clustered in the "saline" (cells in homeostatic condition) and in the " main LPS" (inflammatory condition). By a visual investigation on the dimension-reduced t-SNE space reported in Fig.**4A** of the corresponding publication below, an LPS subset assembled closer to the control condition highlighting the presence of a specific microglia subpopulation under inflammatory conditions. Since the spatial vicinity indicated transcriptional similarity, the LPS subgroup seemed to correspond to a less activated state associated to either less sensitive cells to inflammatory stimuli or a cluster of already recovered cells following the acute pro-inflammatory response.

To investigate in detail the gene expression changes characterizing this "subset LPS", I applied branching analysis modelling (BEAM) and the corresponding pseudotime analysis implemented in Monocle 2 [138]. This approach was able to distinguish between 9 sub-states over the global "main LPS" cluster. The previous identified LPS subpopulation corresponded to the yellow state with a core of cells indicating a specific phenotype characterization that was lost progressively on the branch towards to more indicative inflammatory states (Fig.**5A** in the paper below). The BEAM approach is a generalized regression modelling, which I applied to single-cell data to uncover transcriptional pathways activated within the different subpopulations. The related pseudotime quantity was then referred to the single-cell asynchronous response during inflammation, which has indicated that the "LPS subgroup" correspond to a delayed activation state of microglia. Finally I analyzed gene-expression profile changes for some inflammatory and homeostatic genes of cells in "LPS main" clustered ordered in respect to increasing pseudotime direction. This comparison has characterized the intermediate state where the gene expression levels of the inflammatory genes are increased while the homeostatic gene markers are not yet unexpressed (Fig.**5C**), confirming finally the

unexpected molecular variability in the microglia activation states.

4.2   Paper related to heterogeneity of microglia signatures

Scientific Report

SOURCE
DATA

TRANSPARENT
PROCESS

OPEN
ACCESS

EMBO
reports

Single-cell transcriptomics reveals distinct inflammation-induced microglia signatures

Carole Sousa[1,2,3], Anna Golebiewska[1], Suresh K Poovathingal[2,4], Tony Kaoma[5], Yolanda Pires-Afonso[1,3], Silvia Martina[2], Djalil Coowar[2], Francisco Azuaje[5], Alexander Skupin[2,6], Rudi Balling[2], Knut Biber[7,8] (iD), Simone P Niclou[1,9] (iD) & Alessandro Michelucci[1,2,*] (iD)

Abstract

Microglia are specialized parenchymal-resident phagocytes of the central nervous system (CNS) that actively support, defend and modulate the neural environment. Dysfunctional microglial responses are thought to worsen CNS diseases; nevertheless, their impact during neuroinflammatory processes remains largely obscure. Here, using a combination of single-cell RNA sequencing and multicolour flow cytometry, we comprehensively profile microglia in the brain of lipopolysaccharide (LPS)-injected mice. By excluding the contribution of other immune CNS-resident and peripheral cells, we show that microglia isolated from LPS-injected mice display a global downregulation of their homeostatic signature together with an upregulation of inflammatory genes. Notably, we identify distinct microglial activated profiles under inflammatory conditions, which greatly differ from neurodegenerative disease-associated profiles. These results provide insights into microglial heterogeneity and establish a resource for the identification of specific phenotypes in CNS disorders, such as neuroinflammatory and neurodegenerative diseases.

Introduction

The healthy brain hosts distinct and specialized populations of tissue-resident macrophages strategically placed in the parenchyma, perivascular spaces, meninges and choroid plexus where they coordinate homeostatic and immune surveillance functions [1]. As the only parenchymal-resident immune cells of the central nervous system (CNS), microglia act as critical effectors and regulators of changes in the CNS during development and adult homeostasis. Their ontogeny, together with the absence of turnover from the periphery and the exceptional environment of the CNS, makes microglia a unique immune cell population [2]. By sensing any disruption of CNS homeostasis, microglia rapidly change their gene expression programmes and functional profiles. Recent genome-wide transcriptional studies revealed a unique molecular signature selectively expressed in homeostatic microglia [3–6] that is lost in disease and during ageing [4,7–17]. Microglia coordinate immune responses between the periphery and the CNS as they perceive and propagate inflammatory signals initiated outside the CNS [18]. A multitude of signals received from the CNS environment as well as from the periphery induce microglial responses towards phenotypes that ultimately may support or harm neuronal health [2,19]. Although neuroinflammation and its associated immune responses are often linked to neurodegeneration, the inflammatory response *per se* provides a primary, transient and self-limiting defence mechanism, by which harmful stimuli are resolved and tissue damage is repaired [20]. Disruption of CNS homeostasis, neuronal deterioration and inflammation are common pathophysiological features of several neurodegenerative diseases. In this context, chronic inflammation is likely to be triggered by abnormal protein deposition, by signals elicited by injured neurons and synapses or by impaired pro- and anti-inflammatory regulatory mechanisms that ultimately exacerbate the neurodegenerative process [21]. Dysfunctional microglial responses are believed to worsen CNS diseases [22]; nevertheless, their impact during the neuroinflammatory processes remains largely obscure.

1   NORLUX Neuro-Oncology Laboratory, Department of Oncology, Luxembourg Institute of Health, Luxembourg, Luxembourg
2   Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Esch-Belval, Luxembourg
3   Doctoral School of Science and Technology, University of Luxembourg, Esch-sur-Alzette, Luxembourg
4   Single Cell Analytics & Microfluidics Core, Vlaams Instituut voor Biotechnologie-KU Leuven, Leuven, Belgium
5   Proteome and Genome Research Unit, Department of Oncology, Luxembourg Institute of Health, Luxembourg, Luxembourg
6   National Centre for Microscopy and Imaging Research, University of California San Diego, La Jolla, CA, USA
7   Section Molecular Psychiatry, Department for Psychiatry and Psychotherapy, Laboratory of Translational Psychiatry, Medical Center - University of Freiburg, Faculty of Medicine, University of Freiburg, Freiburg, Germany
8   Section Medical Physiology, Department of Neuroscience, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands
9   Department of Biomedicine, KG Jebsen Brain Tumour Research Center, University of Bergen, Bergen, Norway
    *Corresponding author. Tel: +352 26970 263; E-mail: alessandro.michelucci@lih.lu

In recent years, single-cell RNA sequencing investigations have emerged as a remarkable method to depict heterogeneous cell populations and measure cell-to-cell expression variability of thousands of genes [23–25]. In the murine and human brains, single-cell RNA sequencing analyses have revealed neural and glial cell heterogeneity [26–30]. Similarly, the complexity of immune cell types has been recently unravelled [31]. However, although recent studies have elucidated microglia signatures associated with inflammatory conditions at the bulk level [4,16,32], it is still not clear whether all microglial cells uniformly react to the inflammatory stimuli.

To elucidate the heterogeneity of microglial responses towards systemic inflammation, we here analysed the effect of a peripheral injection of the Gram-negative bacterial endotoxin lipopolysaccharide (LPS) in 3- to 4-month-old C57BL/6N mice using a combination of multicolour flow cytometry and single-cell RNA sequencing analyses. LPS is a well-known immunostimulant used to mimic inflammatory and infectious conditions inducing immune responses associated with sickness behaviour in mice and humans [33,34]. Notably, it has been shown that repeated peripheral injections of LPS in mice induce neurodegeneration, while a single-dose injection of LPS induces acute inflammatory, but not neurodegenerative processes [35]. By our approach, we have identified distinct microglial activated profiles under acute inflammatory conditions, which differ from the recently described disease-associated phenotypes [14].

Understanding the specific molecular triggers and the subsequent genetic programmes defining microglia under homeostatic, inflammatory and neurodegenerative conditions at the single-cell level is a fundamental step to further uncover the multifaceted nature of microglia, thus opening new windows to design novel therapeutic strategies to restore, for example, efficient inflammatory immune responses in CNS diseases.

Results and Discussion

Acutely isolated CD11b$^+$CD45^{int} cells express high levels of microglial homeostatic genes and represent a specific resident immune cell population

Cell-specific transcriptomic analyses are critically dependent on isolation protocols to obtain pure populations resembling their physiological profiles. To characterize microglia close to their proper environment, mouse brains were mechanically dissociated into single-cell suspension with all the steps performed at 4°C [36]. Since microglia in the mouse brain represent only 10% of the cells, CD11b$^+$CD45^{int} microglia were purified from other CNS and immune cells, including CD11b$^+$CD45^{high} macrophages and CD11b$^-$CD45^{high} lymphocytes, by FACS, as described previously (Figs 1A and EV1) [37]. To verify accurate microglial enrichment, we compared gene expression levels of specific CNS cell type markers between RNA extracted from unsorted total brain cells and CD11b$^+$CD45^{int} sorted microglia (Fig 1B). We analysed the expression levels of microglial homeostatic genes (*Olfml3*, *Fcrls*, *Tmem119*, *Siglech*, *Gpr34*, *P2ry12*) as well as astrocytic (*Gfap*, *Gjb6*, *Ntsr2*, *Aldh1l1*), oligodendrocytic (*Mobp*, *Mog*, *Cldn11*) and neuronal (*Tubb3*, *Vglut1*, *NeuN*) markers. As expected, microglial markers were highly expressed in CD11b$^+$CD45^{int} sorted cells, whereas astrocytic, oligodendrocytic and neuronal markers were undetectable or detectable at background levels (Figs 1B and EV1). We next investigated whether CD11b$^+$CD45^{int} population contained resident non-parenchymal macrophages, such as perivascular macrophages. This was inferred using CD206 as an additional marker for resident macrophages [38]. Under homeostatic conditions, CD11b$^+$CD45^{int} microglia contained only $0.04 \pm 0.02\%$ CD206$^+$ cells, while CD11b$^+$CD45^{high} cells contained $24.7 \pm 3.8\%$ CD206$^+$ resident macrophages (Fig 1C and D). Similar results were obtained for the dendritic cell marker CD11c and the monocytic markers Ly6C and CCR2 (Fig EV1). Taken together, these results show that our approach highly discriminates pure and not activated microglial populations from other resident CNS cells.

Microglia isolated from LPS-injected mice show a classical activated pro-inflammatory profile accompanied by a decreased homeostatic signature

The response of microglia towards specific pro- or anti-inflammatory cues *in vitro* has been extensively studied [39]. Treatment of primary microglial cells with TGF-β, LPS or IL-4 generates, respectively, the so-called M0 homeostatic, M1 pro-inflammatory and M2 anti-inflammatory states defined by specific gene signatures [5,40]. However, our understanding towards the reaction of microglia under inflammatory conditions *in vivo* is only starting to emerge. To comprehensively investigate the effect of a systemic inflammatory and/or infectious state on microglia, we peripherally injected mice with LPS (4 μg/g body) 24 h prior analysis. It has been shown that a single-dose injection of LPS induces acute inflammatory, but not neurodegenerative processes [35]. We isolated CD11b$^+$CD45^{int} cells from LPS-injected mice and compared mRNA levels of specific genes

Figure 1. Characterization of acutely isolated CD11b$^+$CD45^{int} cells.

A   FACS gating strategy representative of five independent experiments adopted to sort CD11b$^+$CD45^{int} microglia distinctly from CD11b$^+$CD45^{high} resident macrophages and CD11b$^-$CD45^{high} lymphocytes.

B   Analysis of relative transcript levels of CD11b$^+$CD45^{int} FACS-sorted microglia compared with whole brain tissue by qPCR. Gene expression levels of microglia (*Olfml3*, *Fcrls*, *Tmem119*, *Siglech*, *Gpr34*, *P2ry12*), astrocyte (*Gfap*, *Gjb6*, *Ntsr2*, *Aldh1l1*), oligodendrocyte (*Mobp*, *Mog*, *Cldn1*) and neuron (*Tubb3*, *Vglut1*, *NeuN*) markers. Bars represent mean (n = 4; pool of one female and one male per biological replicate) of relative expression (*Gapdh* as housekeeping gene) $\pm$ SEM (*$P < 0.05$; **$P < 0.01$ by two-tailed Student's *t*-test). N.D., not detected.

C   Representative quantification of CD206 expression in CD11b$^+$CD45^{int} microglia and CD11b$^+$CD45^{high} resident macrophages. Values denote the percentage of the mean $\pm$ SEM of five independent experiments.

D   Representative images of two independent experiments showing microglia, resident macrophages and lymphocytes acquired with ImageStream imaging cytometer (Amnis) based on CD45, CD11b and CD206 expression levels (scale bar represents 7 μm).
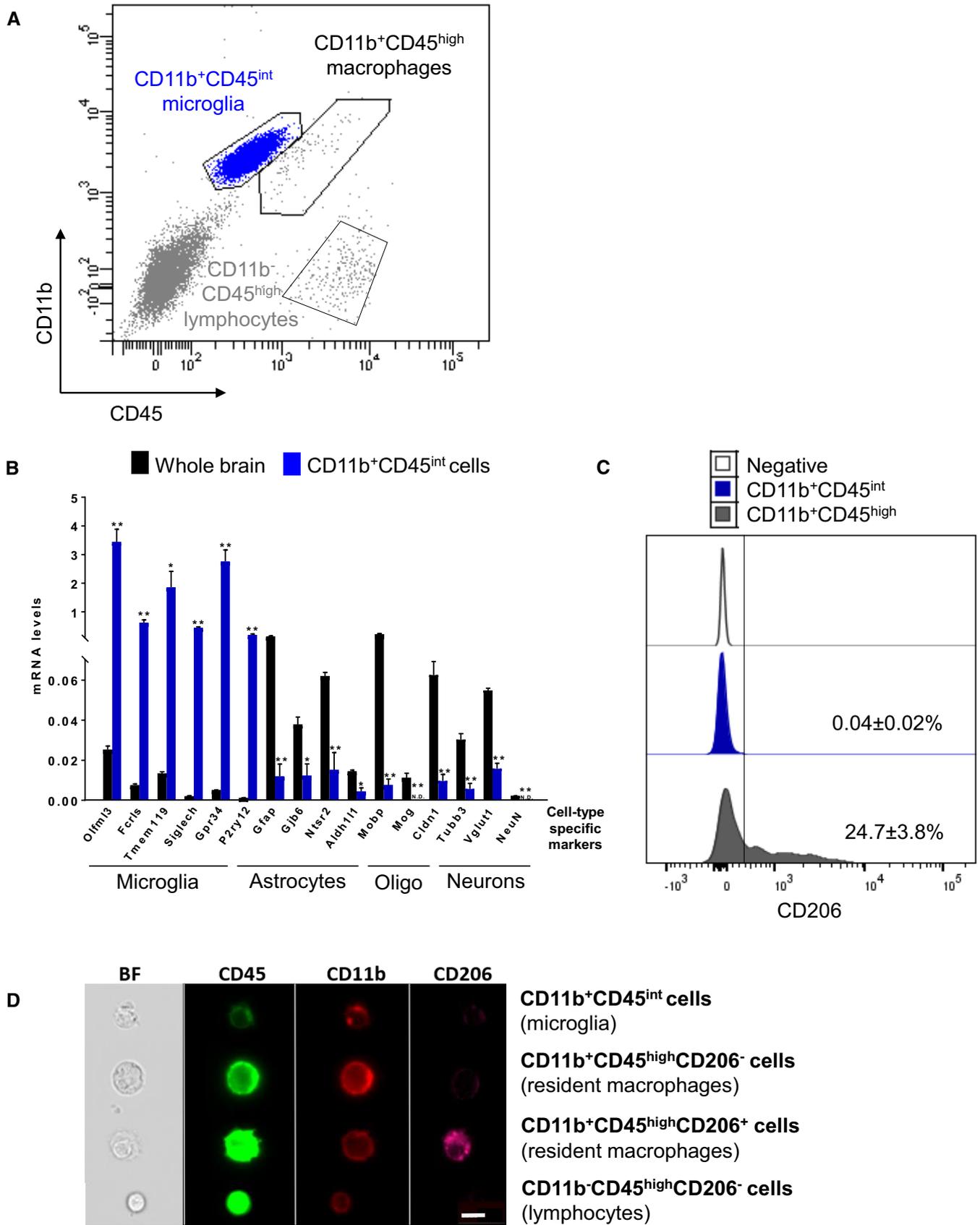
Figure 1.

to the corresponding cells isolated from saline-injected control mice by qPCR. In agreement with previous studies [32,41], the expression levels of homeostatic (e.g. *Olfml3, Fcrls, Tmem119, Siglech, Gpr34, P2ry12, Mef2c*), phagocytic (*Tyrobp* and *Trem2*) and anti-inflammatory genes (e.g. *Mrc1* and *Arg1*) were highly decreased in microglia isolated from LPS-injected mice compared to untreated mice, while the classical pro-inflammatory genes (e.g. *Il1b, Tnf* and *Ccl2*) were markedly increased (Figs 2A and EV2). Notably, it has been recently shown that signals from the CNS microenvironment have considerable influence in shaping, maintaining and reinforcing microglial identity by regulating expression and establishing distinct chromatin landscapes surrounding enhancer regions [42–44]. Changes in chromatin remodelers associate with changes in the expression of nearby genes. Specifically, MEF2C binding sites were shown to be over-enriched in enhancer regions of microglial-specific genes [42] and the loss of MEF2C was associated with priming of microglia [45]. In line with these observations, *Mef2c* expression levels were highly decreased in microglia isolated from LPS-injected mice compared to naïve mice.

We verified that this signature is microglia-specific, and it is not affected by LPS-activated immune peripheral cells, such as lymphocytes (CD11b⁻CD45^high cells) and peripheral monocytes/macrophages (CD11b⁺CD45^high cells), as no significant differences were detected between cellular populations present in brains of saline- and LPS-injected mice (Figs 2B and EV2). Importantly, CD11b⁺CD45^int FACS-gated cells contained very rare (< 0.25%) Ly6C⁺ putative monocytes and (< 0.1%) CD206⁺ putative resident macrophages (Fig 2C). Also, the expression of monocytic markers *Ly6c1* and *Ccr2* was very low in CD11b⁺CD45^int microglia compared to bone marrow-isolated monocytes with no significant differences under LPS exposure (Figs 2D and EV2). In order to further assess that the decrease in the homeostatic signature under inflammatory conditions is not due to the presence of other immune cell types, but it is an intrinsic property of microglial cells, we also analysed the effect of LPS on cultivated microglial from adult and neonatal mice. As expected, the expression level of the homeostatic genes was markedly decreased in cultivated cells when compared to acutely isolated microglia (Fig EV2) [5]. Thus, we cultivated adult microglia in the presence of TGF-β (50 μg/ml) and M-CSF (10 ng/ml) or neonatal cells with TGF-β 24 h prior treatment with LPS to induce the expression of the homeostatic genes, although at a lower extent than in *ex vivo* isolated cells (Fig EV2). Cells treated with LPS (1 ng/ml) for 6 h showed a dramatic decrease of the expression

levels of the homeostatic gene markers, such as *Olfml3, Tmem119* and *Gpr34*, accompanied by enhanced expression levels of inflammatory marker genes, such as *Il1b, Tnf* and *Ccl2* both in adult and in neonatal microglia when compared to cells treated with TGF-β only (Fig 2E). In the healthy brain, TGF-β is expressed at low levels by both neurons and glial cells [46,47], while its expression is increased upon injury [48,49], hypoxia–ischaemia [50] and neurodegeneration [51,52]. SMAD and signal transducer and activator of transcription (STAT) proteins are key signal transducers and transcription factors controlling TGF-β downstream signalling [53]. Specifically, STAT3 and suppressor of cytokine signalling 3 (SOCS3) regulate inflammatory responses [54]. The binding of SOCS3 to both JAK kinase and the cytokine receptor results in the inhibition of STAT3 activation. In our analysis, microglial cells treated with LPS showed increased amounts of STAT3 phosphorylation along with upregulation of *Socs3* expression levels compared to untreated cells (Appendix Fig S1). Taking advantage of the "harmonizome" collection of databases [55], we attested that more than 1/3 of the top 100 sensome genes [4] possess STAT3-binding sites in their promoter region. Hence, we hypothesized that the SOCS3-STAT3 antagonistic signalling may be responsible for the suppression of the homeostatic microglia signature and the concomitant shift towards the inflammatory profile [56].

These results show that microglia isolated from LPS-injected mice display a classical activated pro-inflammatory profile associated with a decrease in the expression of the homeostatic genes. The decrease in the homeostatic signature under inflammatory conditions is an inherent facet of microglial *in vivo* and *in vitro*.

Single-cell mRNA sequencing of CD11b⁺CD45^int microglia isolated from LPS-injected mice reveals a global transcriptional shift and increased heterogeneity compared to steady state conditions

Based on the observed differences in the targeted qPCR approach under steady state and LPS conditions, we next aimed to investigate microglial states at the genome-wide level and infer their transcriptomic heterogeneity at single-cell resolution, since studying a population of cells masks the differences among individual cells. For this purpose, FACS-sorted CD11b⁺CD45^int cells from saline- or LPS-injected mice were analysed using the recently developed high-throughput droplet-based Drop-seq method [23]. In Drop-seq, single cells and functionalized barcoded beads as cell identifiers are co-encapsulated into droplets followed by cDNA synthesis,

Figure 2. LPS stimulation induces an intrinsic loss of the microglia homeostatic signature.

A–D   Three- to four-month-old C57BL/6N mice were treated with an acute dose of LPS (4 μg/g body) or vehicle (saline). Microglia (pool of two mice per group per replicate; one female and one male) were FACS-sorted 24 h later. (A) Gene expression levels of microglial homeostatic (*Olfml3, Fcrls, Tmem119, Siglech, Gpr34, P2ry12, Mef2c*), phagocytic (*Tyrobp, Trem2*) and inflammatory (*Il1b, Tnf, Ccl2, Mrc1, Arg1*) markers were analysed by qPCR. Bars represent mean of relative expression (% of saline; *Gapdh* as housekeeping gene) ± SEM (*P < 0.05; **P < 0.01 by two-tailed Student's *t*-test; *n* = 4). (B) Representative multicolour flow cytometry analysis of five independent experiments showing CD11b- and CD45-positive populations in single viable cells in saline or LPS-injected mouse brains. (C) Representative multicolour flow cytometry analysis showing the percentage of the mean ± SEM of five independent experiments of Ly6C- and CD206-expressing cells in CD11b⁺CD45^int cells from saline or LPS-injected mice. (D) Gene expression levels of the monocytic markers *Ly6c1* and *Ccr2* in purified microglia (*n* = 4) and isolated bone marrow monocytes (*n* = 2) by qPCR. Bars represent mean of relative expression (*Gapdh* as housekeeping gene) ± SEM (**P < 0.01 by two-tailed Student's *t*-test).

E   Primary adult microglia were cultivated in the presence of TGF-β (50 μg/ml) and M-CSF (10 ng/ml), while neonatal cells were stimulated for 24 h with TGF-β (50 μg/ml) followed by 6 h of stimulation with LPS (1 ng/ml) or left untreated. Expression levels of microglial homeostatic (*Olfml3, Tmem119, Gpr34*) and inflammatory (*Il1b, Tnf, Ccl2*) genes were analysed by qPCR. Bars represent mean of relative expression (*Gapdh* as housekeeping gene) ± SEM (*P < 0.05; **P < 0.01 by two-tailed Student's *t*-test).
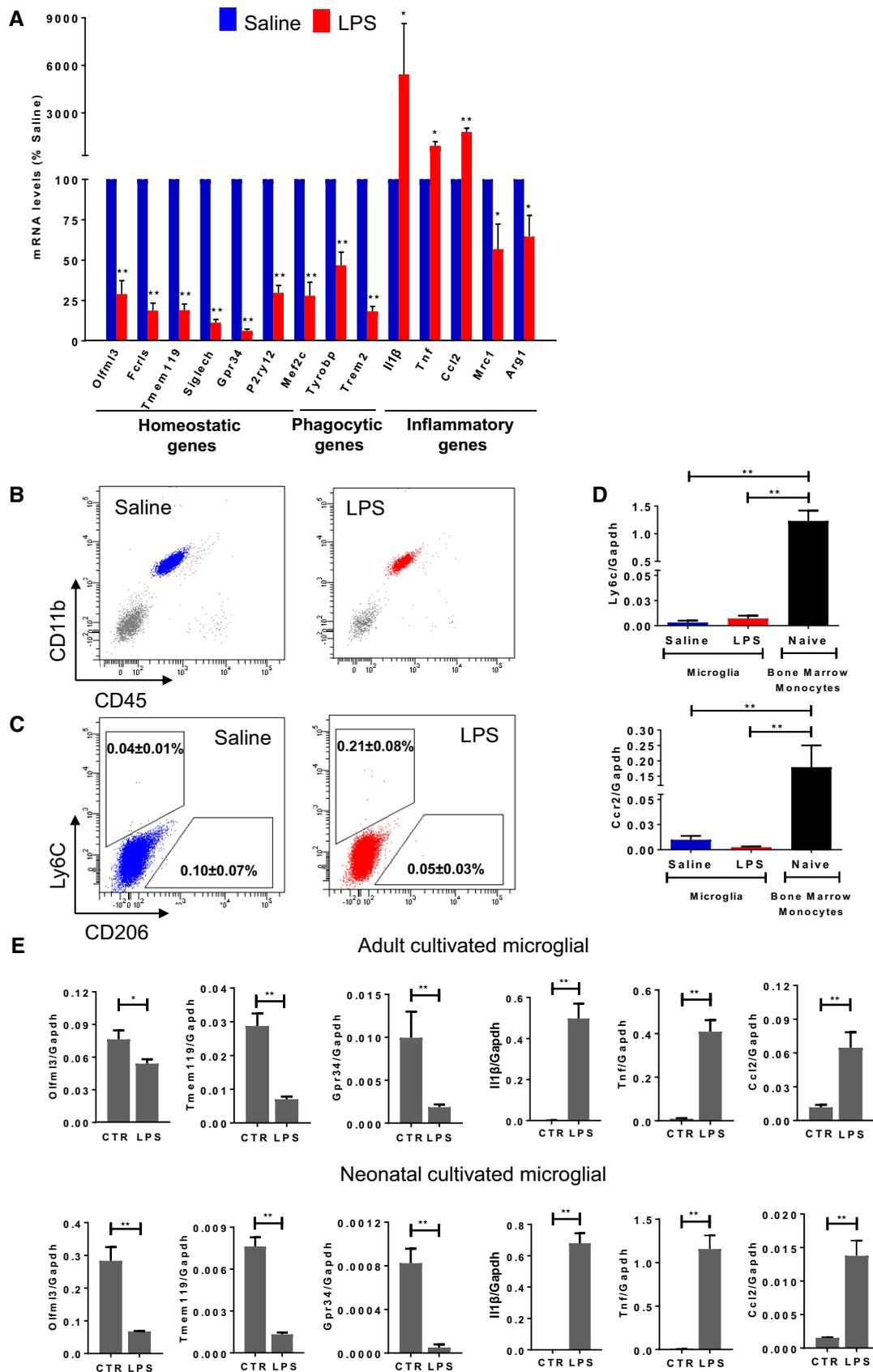
Figure 2.

amplification, library preparation and next-generation sequencing. First, we sought for differentially expressed genes between all LPS and all naïve/saline cells using MAST [57]. We identified 2,405 differentially expressed genes between these two conditions with a false discovery rate (FDR) cut-off of 5% (Dataset EV1) and exemplified the top 100 differentially expressed genes in a heatmap (Fig 3A). Second, principal component analysis followed by two-dimensional *t*-distributed stochastic neighbour embedding (2D-tSNE) of the overall gene expression data of 1,247 analysed cells identified two main cell clusters that were independent of the 2D-tSNE parameters and library sizes (Appendix Fig S2). Microglia isolated from LPS-injected mice distinctly clustered from the corresponding steady state microglia presenting discrete gene expression signatures (Fig 3B; Dataset EV1). Intriguingly, we noticed from both analyses that, although most of the activated cells clustered together, a small group of cells assembled closer to the control cells, thus highlighting the existence of potential subpopulations under inflammatory conditions, which we characterized later. Gene set enrichment analysis (GO) of upregulated genes in microglia isolated from LPS-injected mice using DAVID [58,59] uncovered significant involvement ($P < 2.5 \times 10^{-9}$) in "translation", "protein folding", "ribosome biogenesis" and "immune system process", thus reflecting highly activated cells. On the other hand, GO of the corresponding downregulated genes identified, among others, significant enrichment ($P < 4.9 \times 10^{-5}$) in "regulation of transforming growth factor beta receptor signalling pathway", thus reflecting that TGF-β signalling is among the most affected pathways in microglia exposed to LPS (Appendix Fig S3). In line with gene expression results obtained at the bulk level, microglial homeostatic genes (e.g. *Tmem119, Mef2c, P2ry13, P2ry12, Siglech*) were among the top downregulated genes and classical pro-inflammatory genes (e.g. *Ccl2, Gpr84, Nfkbia*) were mainly upregulated also at the single-cell level (Appendix Fig S2). We further investigated individual gene expressions at single-cell level using 2D-tSNE to show specific homeostatic and inflammatory gene expression levels. For example, *Tmem119, Siglech* and *P2ry12* genes were consistently expressed under steady state, but were downregulated in microglia isolated from LPS-injected mice, while *Ccl2* and *Gpr84* were largely upregulated in most of the cells exposed to LPS compared to saline conditions (Fig 3C; Appendix Fig S4). Notably, a prominent decrease in TMEM119 and P2RY12 expression was further confirmed at the protein level by flow cytometry (Fig 3D).

Although microglial activation is a common hallmark under inflammatory and neurodegenerative conditions [22], microglia transcriptional signatures have been shown to be different. For example, Chiu *et al* [16] demonstrated that acutely isolated microglia from the SOD1^G93A mouse model of amyotrophic lateral sclerosis (ALS) differed from LPS-activated microglia, defining an ALS-specific phenotype. Following the recent description of a novel disease-associated microglial (DAM) phenotype identified under neurodegenerative conditions at single-cell resolution [14], we here compared our inflammatory-associated microglia (IAM) signature to DAM. The scatterplot showing the fold change of genes between microglia isolated from LPS-injected mice (2,405 genes; Dataset EV1) versus DAM (1,660 genes; Dataset EV2) compared to homeostatic microglia (FDR < 0.05) disclosed 1,826 unique genes affected by the LPS treatment (e.g. *Tnf, Irf1*), 1,081 distinct genes in DAM (e.g. *Itgax, Axl*) and 579 shared genes between the two populations

(e.g. *Gpr84, Tmem119*), thus highlighting that these cells mainly display a unique expression profile (Fig EV3). Specifically, only 215 upregulated genes (12.1%) and 364 downregulated genes (21.2%) were shared between the two groups (Fig EV3). Gene set enrichment analysis (GO) and identification of key genes being discriminative between inflammatory microglia and DAM revealed a high inflammatory reactivity upon LPS treatment and a phagocytic/lysosomal gene signature in DAM (Fig EV3). For instances, *Trem2* and *Tyrobp* expression levels were highly decreased in IAM, whereas an elevation of both genes was reported in DAM. TREM2 associates with the immunoreceptor tyrosine-based activation motif (ITAM)-containing adaptor protein TYROBP (DAP12), in which signalling involves the recruitment of tyrosine kinase Syk that further phosphorylates downstream pathways inducing cell activation. TREM2 is required for phagocytosis of apoptotic neurons, microglial proliferation and survival [56,60–62]. These subtle differences in perceiving different signals induced by CNS perturbations support the microglial critical role in modulating specific functional activities. In fact, it is intuitive to consider that sensing inflammatory environments to maintain a homeostatic neuronal network (e.g. through the expression of *Clec4a* and *Clec5a* genes that are exclusively upregulated in our dataset) or recognizing and clearing pathogenic factors (e.g. by expressing *Clec7a/Dectin-1* in DAM), such as β-amyloid aggregates in AD, require distinct activated phenotypes. In a different context, it has been recently shown that myelin pieces are gradually released from ageing myelin sheaths and are subsequently cleared by microglia [63]. Age-related myelin fragmentation is substantial, leading to lysosomal storage and contributing to microglial senescence and immune dysfunction in ageing [63]. It could be then hypothesized that a similar accentuated mechanism may be encountered by microglia surrounding β-amyloid plaques, which become dystrophic at a late stage of the disease [64]. Interestingly, genes described to be associated with neurological diseases, such as *Cd33, Cd9, Sod1, Ctsd*, and *Hif1a*, were also downregulated in our signature in comparison with DAM.

Taken together, these results suggest that microglia under acute systemic inflammation present a highly activated state, which is heterogeneous and distinct from neurodegenerative disease-associated profiles.

Microglia present distinct activated signatures under inflammatory conditions

Next, we aimed to elucidate whether the response to LPS was heterogeneous across microglial cells. Based on our previous observation (Figs 3A and B), we further analysed the identified subclusters by 2D-tSNE representation (Fig 4A). Based on the obtained 2D representation, a specific LPS subgroup ("subset LPS", in yellow) distinct from the core LPS cluster ("main LPS", in red) was identified closer to naïve microglial cluster. Thus, we hypothesized that these cells may correspond to a microglial subset that is less sensitive to inflammatory stimuli or a cluster of cells which already partly recovered from their activated state following the prominent pro-inflammatory immune response. We obtained differentially expressed genes between the "main LPS" (Dataset EV3) and the "subset LPS" (Dataset EV4) clusters compared to the corresponding control conditions (FDR < 0.05). We represented the top 100
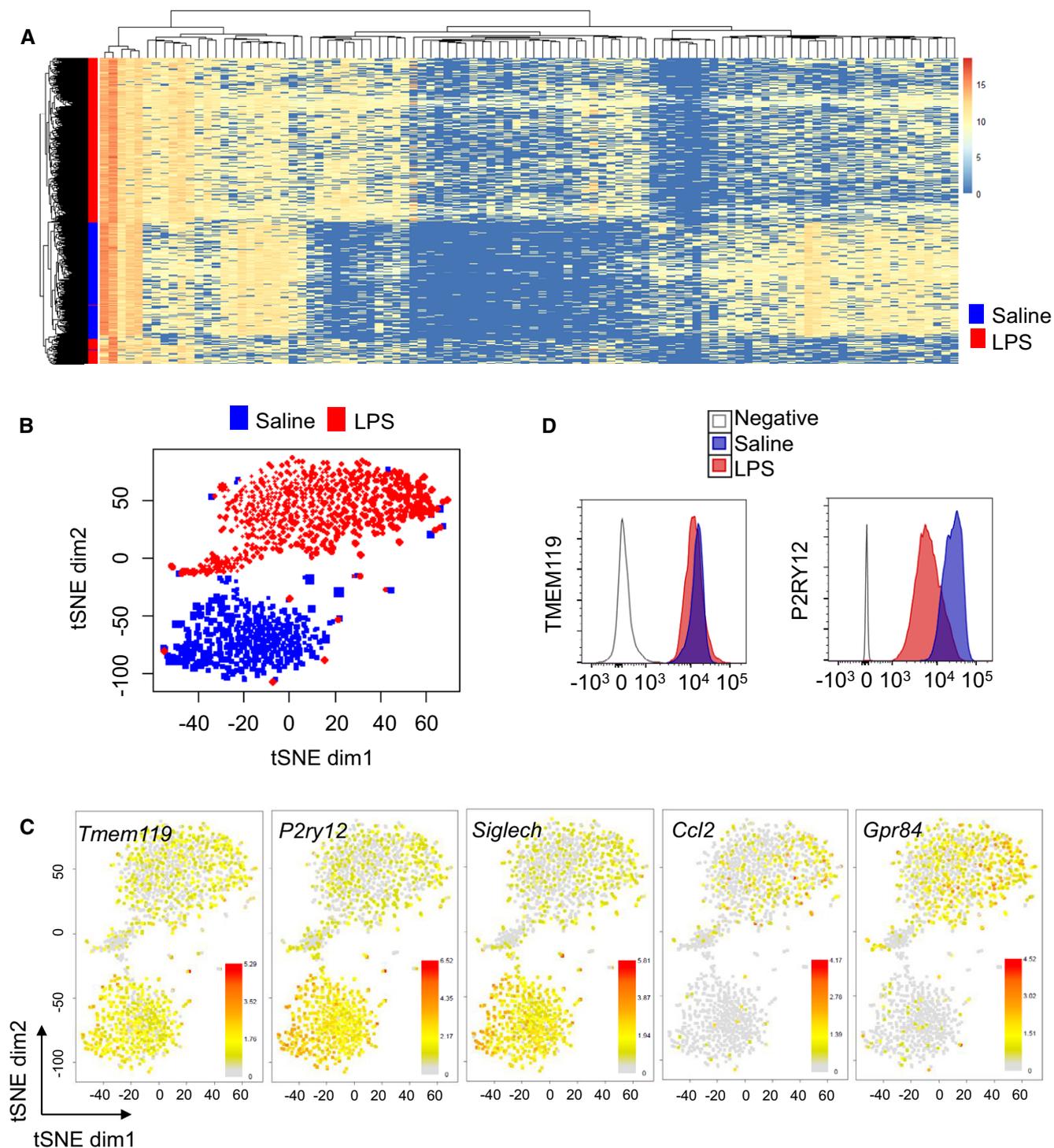
Figure 3. Characterization of microglial activation at the single-cell level.

A   Heatmap showing clustering analysis of 1,247 single cells, featuring 100 most variable genes (FDR < 0.05). Single-cell RNA-seq results are obtained from two mice per group (one female and one male each). Values denote a score based on gene expression rank.

B   2D-tSNE representation of all single cells included in the study (*n* = 1,247) depicting the separation of microglia isolated from LPS-injected mice (770 cells in red) and steady state (477 cells in blue) in two main clusters.

C   Expression of specific homeostatic (*Tmem119, P2ry12, Siglech*) and inflammatory (*Ccl2, Gpr84*) genes overlaid on the 2D-tSNE space. Bars represent log2 (Count + 1).

D   Representative multicolour flow cytometry analysis of two independent experiments showing TMEM119 and P2RY12 expression levels in CD11b$^+$CD45^{int} microglia of saline or LPS-injected mouse brains. For the unconjugated TMEM119 antibody, negative stands for primary antibody without secondary antibody. For P2RY12 antibody, negative represents isotype PE control.
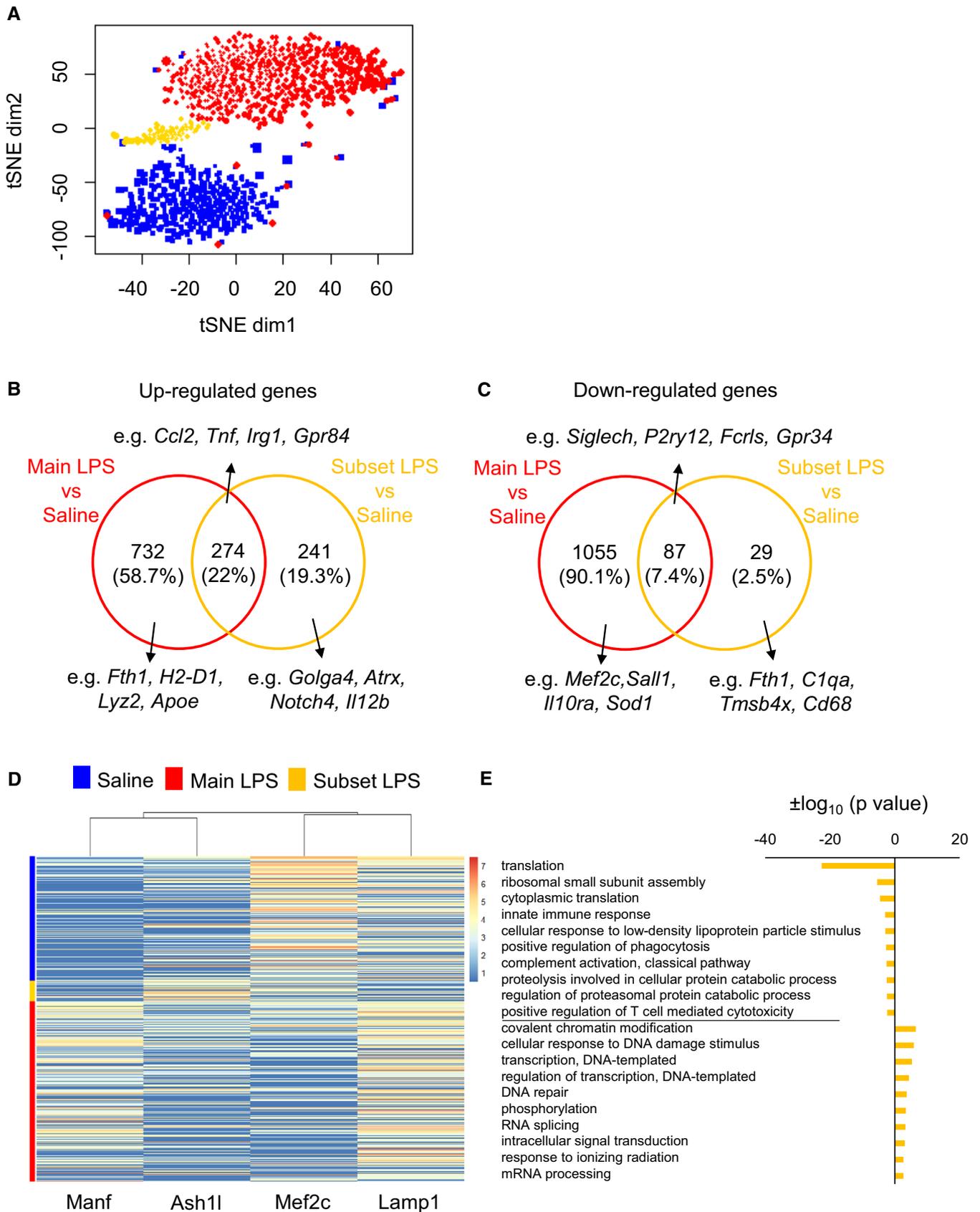
Figure 4.

◄

Figure 4. Identification of microglial subpopulations under inflammatory conditions.

A 2D-tSNE representation of 1,247 single cells isolated from naïve (blue)- and LPS-treated mice showing two distinct subpopulations among the 770 cells isolated from LPS-injected mice (*n* = 703, red; *n* = 67, yellow).

B Venn diagram showing 732 genes uniquely upregulated in the "main LPS" cluster (red) and 241 genes exclusively increased in the "subset LPS" (yellow) compared to their corresponding controls (blue) (FDR < 0.05). A total of 274 genes were shared between the two LPS populations.

C Venn diagram showing 1,055 genes uniquely downregulated in the "main LPS" cluster (red) and 29 genes exclusively decreased in the "subset LPS" (yellow) compared to their corresponding controls (blue) (FDR < 0.05). A total of 87 genes were shared between the two LPS populations.

D Heatmap showing examples of specific genes mainly upregulated in "main LPS" (*Manf*) or "subset LPS" (*Ash1l*) and downregulated in "main LPS" (*Mef2c*) or "subset LPS" (*Lamp1*) overlaid on the 2D-tSNE space. Bars represent log2 (Count + 1).

E Gene set enrichment analysis (GO, top 10 biological processes) of 99 downregulated and 397 upregulated genes distinguishing cells in "subset LPS" versus "main LPS" (FDR < 0.05).

differentially expressed genes among the identified clusters in a heatmap (FDR < 0.05; Appendix Fig S5). To elucidate the transcriptional signature of the LPS subgroups, we showed differentially expressed genes between "main LPS" and "subset LPS" clusters compared to the corresponding control conditions (FDR < 0.05). In line with their activated state, the main pro-inflammatory genes (e.g. *Ccl2, Tnf, Irg1, Gpr84*) were upregulated (Fig 4B) and the microglial homeostatic genes (e.g. *Siglech, P2ry12, Fcrls, Gpr34*) were downregulated in both populations (Fig 4C), although at a lesser extent in "subset LPS", compared to steady state conditions. Investigation of the top differentially expressed genes unique to "main LPS" or "subset LPS" compared to naïve cells (FDR < 0.05; Log2FC ≥ 3 or Log2FC ≤ −3; Table 1) identified, for example, *Manf* (a growth factor that promotes neuroprotection and tissue repair [65]) and *C5ar1* among the top upregulated genes in "main LPS" and *Stab 1* as well as *Ash1l* (which suppresses the production of pro-inflammatory mediators, such as IL-6 and TNF [66]), among the enhanced genes in "subset LPS". Downregulated genes were, for example, the homeostatic gene marker *Mef2c*, which restrains the

Table 1. List of top differentially expressed genes unique to "main LPS" or "subset LPS" versus PBS (FDR < 0.05; upregulated genes, Log2FC ≥ 3; downregulated genes, Log2FC ≤ −3).

Top upregulated genes		Top downregulated genes	
"Main LPS"	"Subset LPS"	"Main LPS"	"Subset LPS"
Rplp0	Gm26924	Tanc2	Lamp1
Rps2	Golga4	Pde3b	Gm17087
Cd52	Zfc3h1	Maf	Cd68
Cd63	RP24-312B12.1	Rasgrp3	Rps14
Ctsl	Stab 1	Zfhx3	C1qc
Manf	Cacna1d	4632428N05Rik	Itm2c
Pdia4	Ash1l	Mef2c	Eif1
Calm1	Ascc3	Qk	H3f3b
Rps19	Atrx	Ivns1abp	Cd81
Fth1	Ptprc	Pmepa1	Ubb
Rps5	Ttc14		Lrrc58
Pdia6	Chd7		
C5ar1	Myo9a		
Ptplb			
Rpl32			
Gnl3			

microglial inflammatory response [45] in "main LPS" and genes associated with endosomes/lysosomes in both "main LPS" (*Maf*) and "subset LPS" (*Lamp1*) (Figs 4D and EV4), thus potentially providing some mechanistic insights regarding the less activated state of "subset LPS" compared to the "main LPS" cluster. Further analysis of unique differentially expressed genes (FDR < 0.05) characterizing the two LPS subpopulations based on microglial functions and properties showed a dramatic increase in genes associated with the major histocompatibility complex (e.g. *H2-D1* and *H2-K1*) exclusively in the "main LPS" group and a decrease in the complement system (e.g. *C1qa, C1qb* and *C1qc*) in the "subset LPS" group when compared to steady state (Table EV1).

Notably, we characterized membrane markers corresponding to specific genes identified at single-cell resolution by flow cytometry to analyse the expression levels of markers upregulated in both LPS groups (e.g. CD44), only in "main LPS" (e.g. CD274) or only in "subset LPS" (e.g. NOTCH4). Although three markers used simultaneously did not allow to clearly discriminate the "subset LPS" from the "main LPS" population, changes in the proportion of marker-positive cells were in line with the scRNA-seq data. Upon LPS treatment, a smaller proportion of NOTCH4-positive cells (saline 5.4%; LPS 18.9%) compared to CD44 (saline 65.2%; LPS 97.5%) and CD274 (saline 48.7%; LPS 88.1%) were detected (Fig EV5). We confirmed this pattern by immunohistochemistry, showing that NOTCH4-positive cells were evenly distributed throughout the brain, thus indicating that these cells were not associated with a specific brain region (Fig EV5).

Gene set enrichment analysis of downregulated genes characterizing "subset LPS" compared to "main LPS" confirmed "innate immune response" and "complement activation, classical pathway" as decreased terms, thus highlighting a less pronounced activated state of the "subset LPS". Intriguingly, these cells revealed significant over-representation (*P* < 0.05) of "covalent chromatin modification" and "DNA repair" that may indicate cells recovering from their acute activated state or a subset of cells with specific chromatin states and DNA repair properties conveying an attenuated activated phenotype than the main population (Fig 4E). In order to further corroborate the existence of the identified microglial subpopulations under inflammatory conditions, the corresponding 770 cells were subjected to the "SC3" method [67]. With two clusters, we found a very high concordance between the subcluster obtained with "SC3" and the LPS subset identified by 2D-tSNE, thus supporting the existence of the detected subpopulations. We represented the top 50 differentially expressed genes driving the segregation of cells into the two clusters in a heatmap (adjusted *P*-value < 0.05; Appendix Fig S6).

A   Branching analysis

B   Pseudotime analysis

C   Pseudotime dynamics of specific genes

Figure 5.

◀

Figure 5. Pseudotime analysis.

A Branching analysis of LPS-activated microglia by Monocle 2 leads to nine distinct clusters in a two-dimensional state space inferred by generalized regression modelling (see Materials and Methods) showing the major difference of "subset LPS" (in yellow) compared to the other clusters corresponding to "main LPS" (in red).

B Monocle estimated a pseudotime for each cell along the inferred cell trajectory within the state space showing a delayed activation pattern of "subset LPS" compared to the other fractions.

C Pseudotime dynamics of inflammatory (*Ccl12, Ccl2, Gpr84*) and homeostatic (*Mef2c, P2ry12, Siglech*) genes in dependence on inferred cell states.

Lastly, we used Venn diagrams to show unique and common upregulated and downregulated (Fig EV5) genes among "main LPS" cluster, "subset LPS" and DAM (FDR < 0.05). Among the deregulated genes, for example, *Spp1*, *Il1b* and *Tlr2* were commonly upregulated, while *Fcrls*, *Tgfbr1* and *Siglech* were downregulated in the three groups. Intriguingly, genes of the complement system (e.g. *C1qa*, *C1qb* and *C1qc*) were downregulated in both "subset LPS" and DAM, but not in "main LPS". Further analysis of the top differentially expressed genes unique to the three groups compared to naïve cells (FDR < 0.05; Log2FC ≥ 3 or Log2FC ≤ −3; Appendix Table S1) showed that the previously identified genes, such as *Manf* and *C5ar1* are uniquely upregulated in "main LPS", while *Stab 1* as well as *Ash1l* is among the increased genes only in "subset LPS".

Overall, these results highlight the existence of specific microglial subpopulations under inflammatory conditions, which are distinct from neurodegenerative-associated phenotypes. These findings emphasize heterogeneity of microglial activated states *in vivo* reflecting specific functional activities related to their corresponding environment.

Pseudotime analysis of LPS-activated microglia uncovers "subset LPS" as an intermediate activated state

Although further analyses at different time points should be performed in the future to resolve the dynamic process of activation, to further investigate the activation process, the heterogeneity within LPS-activated microglia and, specifically, the properties of "subset LPS" compared to "main LPS", we applied branch expression analysis modelling (BEAM) and corresponding pseudotime analysis implemented in Monocle 2 [68]. Since the more subtle differences during the activation process would be dominated by the large differences between naïve and LPS conditions, we applied the branching analysis to the LPS-activated microglia only. This more sensitive analysis revealed nine different states, with the largest difference of "subset LPS" to all others, in accordance with the previous tSNE and "SC3" analyses. Interestingly, cells assigned to "subset LPS" exhibit a dense core, but also a significant number of cells towards the other main clusters (Fig 5A). Given this more sensitive clustering and corresponding pseudotime analysis, we then investigated the characteristics of "subset LPS" with respect to their activation state and their relation with the other states. For this purpose, we plotted the estimated pseudotime of each cell in the state space indicating a delayed activation pattern of "subset LPS" (Fig 5B). Lastly, we investigated inflammatory (e.g. *Ccl12, Ccl2, Gpr84*) and homeostatic (e.g. *Mef2c, P2ry12, Siglech*) gene expression profiles in dependence on pseudotime, further indicating the delayed activation state of "subset LPS" by smaller pseudotimes (Fig 5C). By comparing the dynamics of the inflammatory and homeostatic genes along the activation process, we observed that inflammatory genes were upregulated first, while the homeostatic

markers were only subsequently downregulated. Thus, "subset LPS" may correspond to an intermediate state where the gene expression levels of the inflammatory mediators are increased, but the homeostatic gene markers, such as *Mef2c*, are still partly preserved. In conclusion, from this analysis, we hypothesized that these cells may correspond to a microglial subpopulation that is less sensitive to inflammatory stimuli.

In summary, our work elucidated an extensive picture of microglial profiles in steady state and upon inflammatory conditions, including unforeseen heterogeneity in their states of activation. We believe that our findings, together with the recent single-cell RNA sequencing studies of microglia in Alzheimer's disease [14], present a comprehensive transcriptomic view of microglia under acute inflammatory conditions and a comparison with neurodegenerative processes. These results could then pave the way to design new therapeutic approaches to restore abnormal or detrimental microglial phenotypes found in several CNS disorders.

Materials and Methods

Animals

Three- to four-month-old C57BL/6N male and female mice were obtained from Charles River Laboratories (France). Mice were housed in 12-h light/dark cycle, with sterile food and water *ad libitum*. All animal procedures were approved by the University of Luxembourg Animal Experimentation Ethics Committee and by appropriate government agencies. The animal work of the present study has been conducted and reported in accordance with the ARRIVE (Animal Research: Reporting of *In Vivo* Experiments) guidelines to improve the design, analysis and reporting of research using animals, maximizing information published and minimizing unnecessary studies.

Acute microglial isolation and purification by multicolour flow cytometry

Mice were treated with a single intraperitoneal injection of LPS (4 μg LPS/g body weight) or with PBS (saline) as vehicle control. Twenty-four hours later, mice were deeply anaesthetized with a combination of ketamine (100 mg/ml; Nimatek Vet)–dorbene (medetomidine hydrochloride; 1 mg/ml; Dorbene Vet) and perfused transcardially with ice-cold PBS. Further processing was performed at 4°C and no-break centrifugations. Brains were rapidly removed, stored in ice-cold HBSS (Gibco/Life Technologies) with 1 M HEPES (Gibco/Life Technologies) and 0.5% D-(+)-glucose (Sigma-Aldrich), mechanically homogenized in a potter homogenizer and centrifuged at 900 rpm for 10 min. Myelin was removed from cell suspension with the Myelin Removal Kit (Miltenyi Biotec) according to the manufacturer's protocol. Prior to the FACS, the cell suspension was

resuspended in ice-cold HBSS with 2% FBS and 10 mM HEPES, pH 7.4 and filtered through a 70-μm nylon mesh (CellTrics). For multi-colour staining, cells were incubated for 15 min with Fc receptor binding inhibitor (anti-mouse CD16/CD32 monoclonal antibody; 1:100; eBioscience) to reduce binding of non-specific Fc-gamma receptors, and then stained with fluorochrome-conjugated antibodies (Appendix Table S2) or their corresponding isotopic controls for 45 min at 4°C in dark. After washing, cells were pelleted at 300 g for 10 min at 4°C and resuspended in 200 μl of the appropriated buffer. Hoechst (0.1 μg/ml; Sigma) or Sytox Red (1:1,000; Thermo Fisher Scientific) were added shortly before flow cytometry measurements for dead cell discrimination. Cells were sorted with FACSAria™ SORP cytometer (BD Biosciences) fitted with a 640 nm (30 mW) red laser, a 355 nm (60 mW) UV laser, a 405 nm (50 mW) violet laser, a 488 nm (100 mW) blue laser and a 561 nm (50 mW) yellow/green laser. Data were analysed with FACSDiva software (Becton Dickinson) and FlowJo software (version 7.6.5; Tree Star). Imaging flow cytometry was performed with an ImageStream imaging cytometer (Amnis) fitted with a 375 UV laser, a 488 blue laser, a 561 yellow-green laser, a 642 red laser and a 785 nm infrared laser. Acquisition was performed with the INSPIRE® software, and analysis was performed using IDEAS® image analysis software. Pictures were taken at 60× magnification at low speed, high sensitivity mode.

Isolation of bone marrow monocytes

Monocytes were isolated from mouse bone marrow cells by using the Monocyte Isolation Kit (Miltenyi Biotec) according to the manufacturer's protocol.

Primary adult mouse microglial culture

Adult microglia were isolated from brains of C57BL/6N mice at age 6–10 weeks by magnetic separation. Mice were transcardially perfused with ice-cold PBS under anaesthesia, and brains were dissociated using the Neural Dissociation Kit P (MACS Miltenyi Biotec) according to the manufacturer's instruction. Microglia were enriched by magnetic separation using CD11b$^+$ beads (MACS Miltenyi Biotec). Briefly, 1×10^7 cells were resuspended in 90 μl of MACS buffer [Hank's balanced salt solution (HBSS; Lonza] supplemented with 0.5% BSA (Sigma-Aldrich) and 2 mM EDTA and 10 μl of CD11b MicroBeads (MACS Miltenyi Biotec). The cell suspension was incubated at 4°C for 20 min, washed and pelleted in 500 μl of MACS buffer at a density of 1×10^8 cells. The cell suspension was applied into LS columns (MACS Miltenyi Biotec), and the CD11b$^+$ fraction was eluted. Primary adult microglia were plated in 24-well plates coated in poly-L-lysine (0.1 mg/ml solution; Sigma-Aldrich) at a density of 2×10^5 cells/ml and grown in microglial culture medium [Dulbecco's modified Eagle's medium (DMEM-F12 w/L-glutamine w/15 mM HEPES; Biowest)] supplemented with 10% foetal bovine serum (FBS; Gibco/Life Technologies), pen-strep (100 U/ml/100 μg/ml; Gibco/Life Technologies), human recombinant TGF-β (PeproTech) at a final concentration of 50 μg/ml and mouse recombinant M-CSF (R&D Systems) at a final concentration of 10 ng/ml. Cells were cultured for 5 days without changing media. After 9 days of culture, cells were stimulated with lipopolysaccharide (LPS from *Escherichia coli* 055:B5; Sigma-Aldrich) at a final concentration of 1 ng/ml for 6 h.

Primary newborn mouse microglial culture

Murine primary microglial cells were isolated from newborn (P1–P4) C57BL/6N mouse brains as previously described [69]. Brains were dissected on ice. Subsequently, meninges and large blood vessels were carefully removed and brains were pooled and minced in cold Dulbecco's phosphate buffered saline (PBS; Lonza). Tissue dissociation was completed by 10 min of incubation in 2 mM EDTA (Sigma-Aldrich). Cells were washed, centrifuged, seeded into six-well plates coated with poly-L-lysine and allowed to attach and grow in complete medium DMEM (Gibco/Life Technologies) supplemented with 10% FBS and pen-strep at 37°C in a water-saturated atmosphere containing 5% CO_2. The culture medium was renewed after 3 days of culture. After 10 days, when cells reached confluence, the mixed glial monolayer was trypsinated (0.05% Trypsin–EDTA; Gibco/Life Technologies) and microglial cells were purified by magnetic cell sorting (MACS Miltenyi Biotec) following the manufacturer's instructions. Primary microglia were plated in 12-well plates coated with poly-L-lysine (Sigma-Aldrich) at a density of 4×10^5 cells/ml. Twenty-four hours after plating, cells were activated with different compounds: LPS at a final concentration of 1 ng/ml, TGF-β at a final concentration of 50 μg/ml and M-CSF at a final concentration of 10 ng/ml.

RNA isolation and RT–PCR

CD11b$^+$CD45^{int} cells were FACS-sorted directly to TRIzol® LS, and total RNA was extracted according to the manufacturer's protocol (Life Technologies). RNA from primary cells was extracted using the RNeasy Mini Kit (QIAGEN), according to the manufacturer's instructions. RNA concentration was quantified by NanoDrop (NanoDrop Technologies) and the quality assessed by the quotient of the 28S to 18S ribosomal RNA electropherogram peak using a bioanalyser (Agilent 2100; Agilent Technologies) using a RNA Pico Chip (Agilent Technologies; only samples with RIN ≥ 7 were further analysed). For cDNA synthesis, RNA was reverse-transcribed using SuperScript™ III reverse transcriptase (10,000 U; Invitrogen/Life Technologies) with 1 μl (50 μM)/reaction oligo(dT)20 (25 μM; Invitrogen/Life Technologies) as primer according to the manufacturer's instructions. Reverse transcription was performed at 50°C for 60 min. Gene expression reaction mixtures contained 2 μl of diluted cDNA, 10 μl of Fast SYBR Green Master Mix (Applied Biosystems/Thermo Fisher Scientific) and 0.5 μl of each 10 μM forward and reverse primers. PCRs were carried out in 96-well plates on a ViiA™ 7 real-time PCR system (Applied Biosystems/Thermo Fisher Scientific) using the following programme: 95°C for 20 s, 40 cycles at 95°C for 1 s and 60°C for 20 s. The sequences of the primers designed using Primer-Blast tool are listed in Appendix Table S3. Samples were run in duplicates, and the mean C_t (threshold cycle) values were used to calculate the relative amount of product by the $\Delta\Delta C_t$ method using *Gapdh* as housekeeping gene.

Immunohistochemistry

Under deep ketamine–dorbene anaesthesia, mice were transcardially perfused with ice-cold PBS, post-fixed in 4% paraformaldehyde (PFA) for 48 h and stored at 4°C in 0.02% sodium azide/PBS as preservative. Serialized parasagittal free-floating 50-μm-thick

sections were generated with a vibratome (Leica; VT-1000S) and collected in cryoprotective medium [PBS containing 1-1 ethylene glycol (Sigma-Aldrich) and 1% w/v polyvinylpyrrolidone (Sigma-Aldrich)]. Sections were stored at −20°C in tubes, each containing a series of every 4th section.

For immunofluorescence, a standard protocol was used with minor modifications [70]. Briefly, sections were washed (PBS with 0.1% Triton X-100), permeabilized (PBS with 1.5% Triton X-100), blocked (PBS with 5% BSA) and incubated with primary antibodies (PBS with 0.3% Triton X-100 and 2% BSA): rabbit anti-Iba1 (1:1,000; Wako) and pre-conjugated PE anti-mouse Notch4 (1:80; BioLegend). Iba1 antibody was visualized using goat anti-rabbit IgG Molecular Probes Alexa Fluor 555 (Thermo Fisher Scientific) secondary antibody. Cell nuclei were counterstained with Hoechst (1 µg/ml; Sigma). Sections were mounted on glass slides cover-slipped using Fluoromount™ Aqueous Mounting Medium (Sigma). Microscopic images were obtained using confocal microscopy (Zeiss LSM880).

SDS–PAGE and Western Blotting analysis

Heat-denatured protein samples were separated on 4–12% BisTris–polyacrylamide gel electrophoresis followed by transfer to nitrocellulose membranes 0.2 µm (Bio-Rad). After blocking with 5% (wt/vol) dry milk in TBST for STAT3 and 3% BSA in TBST for Phospho-STAT3, respectively, the membrane was incubated overnight at 4°C in primary anti-STAT3 antibody from mouse (Cell Signaling) diluted 1:1,000 in 5% (wt/vol) dry milk in TBST and in primary anti-Phospho-STAT3 antibody (Cell Signaling) diluted 1:500 in 3% BSA in TBST with constant shaking. After three washing steps with TBS containing 0.1% Tween-20, the membrane was incubated with anti-rabbit antibody or anti-mouse respectively, coupled to horseradish peroxidase and revealed by chemoluminescence using the Pierce™ ECL detection reagents (Thermo Fisher Scientific).

Single-cell RNA sequencing using Drop-seq

Cell preparation
FACS-sorted CD11b[+]CD45[int] cells were collected in pre-cooled HBSS and 0.5% BSA and transferred directly for subsequent Drop-seq analysis. The cells were stored on ice until the start of the Drop-seq experiment (tissue harvest to running of Drop-seq was < 1 h). Prior to cell loading on the Drop-seq chips, the cell viability was verified and the concentration was adjusted to ~150 cells/µl. This was optimal based on Poissonian statistics to achieve single-cell encapsulation within each droplet of ~800–900 pl droplet size. All samples analysed in this work had a cell viability > 95%.

Microfluidics fabrication
Microfluidics devices were generated using a previously published design [23]. Soft lithography was performed using SU-8 2050 photoresist (MicroChem) on 4" silicon substrate to obtain a feature aspect depth of 100 µm. After overnight silanization (using chlorotrimethylsilane; Sigma), the wafer masks were used for microfluidics fabrication. Drop-seq chips were fabricated using silicon-based polymerization chemistry, with the previously published protocol [71]. Briefly, polydimethylsiloxane (PDMS) base and cross-linker (Dow Corning) were mixed at a 10:1 ratio, mixed and

degassed before pouring onto the Drop-seq master template. PDMS was cured on the master template, at 80°C for 2 h. After incubation and cooling, PDMS monoliths were cut and the inlet/outlet ports were punched with 1.25-mm biopsy punchers (World Precision Instruments). The PDMS monolith was plasma-bonded to a clean microscopic glass slide using a Harrick plasma cleaner. Immediately after pairing the plasma-treated surfaces of the PDMS monolith and the glass slide, flow channels of the Drop-seq chip were subjected to a hydrophobicity treatment using 1H,1H,2H,2H-perfluorodecyltri-chlorosilane (in 2% v/v in FC-40 oil; Alfa Aesar/Sigma). After 5 min of treatment, excessive silane was blown through the inlet/outlet ports. Chips were further incubated at 80°C for 15 min.

Single-cell droplet encapsulation
Experiments followed the original Drop-seq protocol [23] with minor changes described below. Synthesized barcoded beads (ChemGenes Corp., USA) were co-encapsulated with cells inside the droplets containing lysis reagents using an optimal bead concentration of 200 beads/µl in Drop-seq Lysis buffer medium. Cellular mRNA was captured on the beads via barcoded oligo (dT) handles synthesized on the surface.

For cell encapsulation, bead suspensions and cell suspension were loaded into 3-ml syringes (BD). To keep beads in homogenous suspension, a micro-stirrer was used (VP Scientific). The QX200 carrier oil (Bio-Rad) used as continuous phase in the droplet generation was loaded into a 20-ml syringe (BD). For droplet generation, 3.6 ml/h and 13 ml/h flowrates were used in KD Scientific Legato Syringe Pumps for the dispersed and continuous phase flows, respectively. After stabilization of droplet formation, the droplet suspension was collected into a 50-ml Falcon tube. Collection of the emulsion was carried out until 1 ml of the single-cell suspension was dispensed. Droplet consistency and stability were evaluated by bright-field microscopy using INCYTO C-Chip Disposable Hemacytometer (Thermo Fisher Scientific). Bead occupancy within droplets was carefully monitored to avoid multiple beads per droplet.

The subsequent steps of droplet breakage, bead harvesting, reverse transcription and exonuclease treatment were carried out in accordance with the Drop-seq method [23]. RT buffer contained 1× Maxima RT buffer, 4% Ficoll PM-400 (Sigma), 1 µM dNTPs (Thermo Fisher Scientific), 1 U/ml RNase Inhibitor (Lucigen), 2.5 µM Template Switch Oligo [23] and 10 U/ml Maxima H-RT (Thermo Fisher Scientific). After Exo-I treatment, the bead counts were estimated using INCYTO C-Chip Disposable Hemacytometer, and 5,000–8,000 beads were aliquoted in 0.2 ml Eppendorf PCR tubes. PCR mix was dispensed in a volume of 50 µl using 1× HiFi HotStart ReadyMix (Kapa Biosystems) and 0.8 mM Template Switch PCR primer. The thermocycling programme for the PCR amplification was modified for the final PCR cycles by 95°C (3 min), four cycles of 98°C (20 s), 65°C (45 s), 72°C (3 min) and 16 cycles of 98°C (20 s), 67°C (20 s), 72°C (3 min), followed by a final extension step of 72°C for 5 min. After PCR amplification, libraries were purified with 0.6× Agencourt AMPure XP beads (Beckman Coulter), according to the manufacturer's protocol. Finally, the purified libraries were eluted in 20 µl RNase/DNase-free Molecular Grade Water. Quality and concentration of the sequencing libraries were assessed using Bioanalyzer High Sensitivity Chip (Agilent Technologies).

NGS preparation for Drop-seq libraries

The 3′ end enriched cDNA libraries were prepared by tagmentation reaction of 600 pg cDNA library using the standard Nextera XT tagmentation kit (Illumina). Reactions were performed according to the manufacturer's instructions. The PCR amplification cycling programme used was 95°C 30 s, and fourteen cycles of 95°C (10 s), 55°C (30 s) and 72°C (30 s), followed by a final extension step of 72°C (5 min). Libraries were purified twice to reduce primers and short DNA fragments with 0.6× and 1× Agencourt AMPure XP beads (Beckman Coulter), respectively, in accordance with the manufacturer's protocol. Finally, purified libraries were eluted in 15 μl Molecular Grade Water. Quality and quantity of the tagmented cDNA library were evaluated using Bioanalyzer High Sensitivity DNA Chip. The average size of the tagmented libraries prior to sequencing was between 400 and 700 bps.

Purified Drop-seq cDNA libraries were sequenced using Illumina NextSeq 500 with the recommended sequencing protocol except for 6pM of custom primer (GCCTGTCCGCGGAAGCAGTGGTATCAACG CAGAGTAC) applied for priming of read 1. Paired-end sequencing was performed for the read 1 of 20 bases (covering the random cell barcode 1–12 bases and the rest 13–20 bases of random unique molecular identifier (UMI) and for read 2 of 50 bases of the genes.

Bioinformatics processing and data analysis

The FASTQ files were assembled from the raw BCL files using Illumina's bcl2fastq converter and ran through the FASTQC codes (Babraham bioinformatics; https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) to check for consistency in library qualities. The monitored quality assessment parameters were (i) quality per base sequence (especially for the read 2 of the gene), (ii) per base N content, (iii) per base sequence content, and (iv) over-represented sequences. Libraries that showed significant deviation were re-sequenced. The FASTQ files were then merged and converted into binaries using PICARD's FastqToSam algorithm. The sequencing reads were converted into a digital gene expression matrix using the Drop-seq bioinformatics pipeline [23].

Data analysis was done in R. Cells with less than 1,000 counts and genes with zero count in all cells were excluded from subsequent analyses, resulting in 1,247 cells (477 from the saline control and 770 from LPS-injected mice) and 12,369 genes. PCA (*prcomp* function *with scaling*) was used for dimensionality reduction, and PCA results were projected onto a two-dimensional (2D) space using *t*-distributed stochastic neighbour embedding (tSNE, *tsne package*, v.0.1-3). As the first principal component was strongly correlated to the total number of UMI (reads) per cell, it was not included in the 2D-tSNE analysis. Differential expression analysis was performed with MAST [57]. *P*-values were adjusted for multiple testing using false discovery rate (FDR) [72]. Prior to MAST analysis, counts were converted into counts per million and log2-transformed. For subpopulation identification, two approaches were used: (i) based on visual inspection of 2D-tSNE plot, cells were divided into three clusters: one cluster contained almost exclusively cells isolated from control mice, another cluster contained mainly cells harvested from LPS-injected mice, and the last cluster was constituted of a small subset of LPS-derived cells. Clusters were pruned to keep only cells coming from the predominant sample in the group. Comparisons of gene expression between different groups were done with the Kruskal–Wallis *H*-test. *P*-values were corrected with FDR [72]; (ii) each condition was analysed separately

with the "SC3" package [67]. Branching analysis was performed by Monocle 2.4.0 in R (version 3.4.4) with standard parameters [68,73]. The branching method orders cells along an estimated cell trajectory within a gene expression state space based on gene expression similarities estimated by generalized linear regression models.

Statistical analysis

Statistical analyses for qPCRs and FACS experiments were performed using GraphPad Prism 7 software. Comparisons of two groups were performed with a two-tailed Student's *t*-test. Comparisons involving more than two groups were performed using one-way ANOVA followed by the Bonferroni correction for multiple testing. All differences were considered significantly different at $P < 0.05$. Further statistical analysis details are reported in the figure legends.

Data availability

Single-cell RNA sequencing data have been deposited in Gene Expression Omnibus (GEO) database under the accession number GSE115571 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc = GSE115571).

Expanded View for this article is available online.

Author contributions

CS and AM designed the project; KB, RB and SPN involved in the experimental design; CS, SKP, YP-A and AM performed experiments; CS, AG, SKP, TK, SM, AS and AM analysed experiments; DC provided animals; FA supervised the bioinformatics analyses of the single-cell RNA-seq; CS and AM wrote the manuscript; and AG, SKP, TK, DC, YP-A, SM, FA, AS, RB, KB and SPN edited and approved the manuscript.

Conflict of interest

The authors declare that they have no conflict of interest.

References

1. Prinz M, Priller J (2014) Microglia and brain macrophages in the molecular age: from origin to neuropsychiatric disease. *Nat Rev Neurosci* 15: 300–312

2. Sousa C, Biber K, Michelucci A (2017) Cellular and molecular characterization of microglia: a unique immune cell population. *Front Immunol* 8: 198

3. Gautier EL, Shay T, Miller J, Greter M, Jakubzick C, Ivanov S, Helft J, Chow A, Elpek KG, Gordonov S *et al* (2012) Gene-expression profiles and transcriptional regulatory pathways that underlie the identity and diversity of mouse tissue macrophages. *Nat Immunol* 13: 1118 – 1128

4. Hickman SE, Kingery ND, Ohsumi TK, Borowsky ML, Wang LC, Means TK, El Khoury J (2013) The microglial sensome revealed by direct RNA sequencing. *Nat Neurosci* 16: 1896 – 1905

5. Butovsky O, Jedrychowski MP, Moore CS, Cialic R, Lanser AJ, Gabriely G, Koeglsperger T, Dake B, Wu PM, Doykan CE *et al* (2014) Identification of a unique TGF-beta-dependent molecular and functional signature in microglia. *Nat Neurosci* 17: 131 – 143

6. Zhang Y, Chen K, Sloan SA, Bennett ML, Scholze AR, O'Keeffe S, Phatnani HP, Guarnieri P, Caneda C, Ruderisch N *et al* (2014) An RNA-sequencing transcriptome and splicing database of glia, neurons, and vascular cells of the cerebral cortex. *J Neurosci* 34: 11929 – 11947

7. Butovsky O, Jedrychowski MP, Cialic R, Krasemann S, Murugaiyan G, Fanek Z, Greco DJ, Wu PM, Doykan CE, Kiner O *et al* (2015) Targeting miR-155 restores abnormal microglia and attenuates disease in SOD1 mice. *Ann Neurol* 77: 75 – 99

8. Elliott R, Li F, Dragomir I, Chua MM, Gregory BD, Weiss SR (2013) Analysis of the host transcriptome from demyelinating spinal cord of murine coronavirus-infected mice. *PLoS One* 8: e75346

9. Holtman IR, Raj DD, Miller JA, Schaafsma W, Yin Z, Brouwer N, Wes PD, Moller T, Orre M, Kamphuis W *et al* (2015) Induction of a common microglia gene expression signature by aging and neurodegenerative conditions: a co-expression meta-analysis. *Acta Neuropathol Commun* 3: 31

10. Lewis ND, Hill JD, Juchem KW, Stefanopoulos DE, Modis LK (2014) RNA sequencing of microglia and monocyte-derived macrophages from mice with experimental autoimmune encephalomyelitis illustrates a changing phenotype with disease course. *J Neuroimmunol* 277: 26 – 38

11. Olah M, Amor S, Brouwer N, Vinet J, Eggen B, Biber K, Boddeke HW (2012) Identification of a microglia phenotype supportive of remyelination. *Glia* 60: 306 – 321

12. Verheijden S, Beckers L, Casazza A, Butovsky O, Mazzone M, Baes M (2015) Identification of a chronic non-neurodegenerative microglia activation state in a mouse model of peroxisomal beta-oxidation deficiency. *Glia* 63: 1606 – 1620

13. Wang Y, Cella M, Mallinson K, Ulrich JD, Young KL, Robinette ML, Gilfillan S, Krishnan GM, Sudhakar S, Zinselmeyer BH *et al* (2015) TREM2 lipid sensing sustains the microglial response in an Alzheimer's disease model. *Cell* 160: 1061 – 1071

14. Keren-Shaul H, Spinrad A, Weiner A, Matcovitch-Natan O, Dvir-Szternfeld R, Ulland TK, David E, Baruch K, Lara-Astiaso D, Toth B *et al* (2017) A unique microglia type associated with restricting development of Alzheimer's disease. *Cell* 169: 1276 – 1290 e17

15. Krasemann S, Madore C, Cialic R, Baufeld C, Calcagno N, El Fatimy R, Beckers L, O'Loughlin E, Xu Y, Fanek Z *et al* (2017) The TREM2-APOE pathway drives the transcriptional phenotype of dysfunctional microglia in neurodegenerative diseases. *Immunity* 47: 566 – 581 e9

16. Chiu IM, Morimoto ET, Goodarzi H, Liao JT, O'Keeffe S, Phatnani HP, Muratet M, Carroll MC, Levy S, Tavazoie S *et al* (2013) A neurodegeneration-specific gene-expression signature of acutely isolated microglia from an amyotrophic lateral sclerosis mouse model. *Cell Rep* 4: 385 – 401

17. Orre M, Kamphuis W, Osborn LM, Jansen AHP, Kooijman L, Bossers K, Hol EM (2014) Isolation of glia from Alzheimer's mice reveals inflammation and dysfunction. *Neurobiol Aging* 35: 2746 – 2760

18. Norden DM, Godbout JP (2013) Review: microglia of the aged brain: primed to be activated and resistant to regulation. *Neuropathol Appl Neurobiol* 39: 19 – 34

19. Crotti A, Ransohoff RM (2016) Microglial physiology and pathophysiology: insights from genome-wide transcriptional profiling. *Immunity* 44: 505 – 515

20. Fullerton JN, Gilroy DW (2016) Resolution of inflammation: a new therapeutic frontier. *Nat Rev Drug Discov* 15: 551 – 567

21. Wyss-Coray T, Mucke L (2002) Inflammation in neurodegenerative disease–a double-edged sword. *Neuron* 35: 419 – 432

22. Glass CK, Saijo K, Winner B, Marchetto MC, Gage FH (2010) Mechanisms underlying inflammation in neurodegeneration. *Cell* 140: 918 – 934

23. Macosko EZ, Basu A, Satija R, Nemesh J, Shekhar K, Goldman M, Tirosh I, Bialas AR, Kamitaki N, Martersteck EM *et al* (2015) Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* 161: 1202 – 1214

24. Shalek AK, Satija R, Adiconis X, Gertner RS, Gaublomme JT, Raychowdhury R, Schwartz S, Yosef N, Malboeuf C, Lu D *et al* (2013) Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells. *Nature* 498: 236 – 240

25. Tang F, Barbacioru C, Wang Y, Nordman E, Lee C, Xu N, Wang X, Bodeau J, Tuch BB, Siddiqui A *et al* (2009) mRNA-Seq whole-transcriptome analysis of a single cell. *Nat Methods* 6: 377 – 382

26. Tasic B, Menon V, Nguyen TN, Kim TK, Jarsky T, Yao Z, Levi B, Gray LT, Sorensen SA, Dolbeare T *et al* (2016) Adult mouse cortical cell taxonomy revealed by single cell transcriptomics. *Nat Neurosci* 19: 335 – 346

27. Mathys H, Adaikkan C, Gao F, Young JZ, Manet E, Hemberg M, De Jager PL, Ransohoff RM, Regev A, Tsai LH (2017) Temporal tracking of microglia activation in neurodegeneration at single-cell resolution. *Cell Rep* 21: 366 – 380

28. Dulken BW, Leeman DS, Boutet SC, Hebestreit K, Brunet A (2017) Single-cell transcriptomic analysis defines heterogeneity and transcriptional dynamics in the adult neural stem cell lineage. *Cell Rep* 18: 777 – 790

29. Chen R, Wu X, Jiang L, Zhang Y (2017) Single-cell RNA-seq reveals hypothalamic cell diversity. *Cell Rep* 18: 3227 – 3241

30. Artegiani B, Lyubimova A, Muraro M, van Es JH, van Oudenaarden A, Clevers H (2017) A single-cell RNA sequencing study reveals cellular and molecular dynamics of the hippocampal neurogenic niche. *Cell Rep* 21: 3271 – 3284

31. Jaitin DA, Weiner A, Yofe I, Lara-Astiaso D, Keren-Shaul H, David E, Salame TM, Tanay A, van Oudenaarden A, Amit I (2016) Dissecting immune circuits by linking CRISPR-pooled screens with single-cell RNA-seq. *Cell* 167: 1883 – 1896 e15

32. Bennett ML, Bennett FC, Liddelow SA, Ajami B, Zamanian JL, Fernhoff NB, Mulinyawe SB, Bohlen CJ, Adil A, Tucker A *et al* (2016) New tools for studying microglia in the mouse and human CNS. *Proc Natl Acad Sci USA* 113: E1738 – E1746

33. Qin L, Wu X, Block ML, Liu Y, Breese GR, Hong JS, Knapp DJ, Crews FT (2007) Systemic LPS causes chronic neuroinflammation and progressive neurodegeneration. *Glia* 55: 453 – 462

34. Dantzer R (2001) Cytokine-induced sickness behavior: mechanisms and implications. *Ann N Y Acad Sci* 933: 222 – 234

35. Bodea LG, Wang Y, Linnartz-Gerlach B, Kopatz J, Sinkkonen L, Musgrove R, Kaoma T, Muller A, Vallar L, Di Monte DA *et al* (2014)

Neurodegeneration by activation of the microglial complement-phago-some pathway. *J Neurosci* 34: 8546–8556

36. Ransohoff RM (2016) A polarizing question: do M1 and M2 microglia exist? *Nat Neurosci* 19: 987–991

37. Sedgwick JD, Schwender S, Imrich H, Dorries R, Butcher GW, ter Meulen V (1991) Isolation and direct characterization of resident microglial cells from the normal and inflamed central nervous system. *Proc Natl Acad Sci USA* 88: 7438–7442

38. Goldmann T, Wieghofer P, Jordao MJ, Prutek F, Hagemeyer N, Frenzel K, Amann L, Staszewski O, Kierdorf K, Krueger M *et al* (2016) Origin, fate and dynamics of macrophages at central nervous system interfaces. *Nat Immunol* 17: 797–805

39. Michelucci A, Heurtaux T, Grandbarbe L, Morga E, Heuschling P (2009) Characterization of the microglial phenotype under specific pro-inflammatory and anti-inflammatory conditions: effects of oligomeric and fibrillar amyloid-beta. *J Neuroimmunol* 210: 3–12

40. Beutner C, Linnartz-Gerlach B, Schmidt SV, Beyer M, Mallmann MR, Staratschek-Jox A, Schultze JL, Neumann H (2013) Unique transcriptome signature of mouse microglia. *Glia* 61: 1429–1442

41. Haynes SE, Hollopeter G, Yang G, Kurpius D, Dailey ME, Gan WB, Julius D (2006) The P2Y12 receptor regulates microglial activation by extracellular nucleotides. *Nat Neurosci* 9: 1512–1519

42. Gosselin D, Link VM, Romanoski CE, Fonseca GJ, Eichenfield DZ, Spann NJ, Stender JD, Chun HB, Garner H, Geissmann F *et al* (2014) Environment drives selection and function of enhancers controlling tissue-specific macrophage identities. *Cell* 159: 1327–1340

43. Lavin Y, Winter D, Blecher-Gonen R, David E, Keren-Shaul H, Merad M, Jung S, Amit I (2014) Tissue-resident macrophage enhancer landscapes are shaped by the local microenvironment. *Cell* 159: 1312–1326

44. Matcovitch-Natan O, Winter DR, Giladi A, Vargas Aguilar S, Spinrad A, Sarrazin S, Ben-Yehuda H, David E, Zelada Gonzalez F, Perrin P *et al* (2016) Microglia development follows a stepwise program to regulate brain homeostasis. *Science* 353: aad8670

45. Deczkowska A, Matcovitch-Natan O, Tsitsou-Kampeli A, Ben-Hamo S, Dvir-Szternfeld R, Spinrad A, Singer O, David E, Winter DR, Smith LK *et al* (2017) Mef2C restrains microglial inflammatory response and is lost in brain ageing in an IFN-I-dependent manner. *Nat Commun* 8: 717

46. Flanders KC, Ludecke G, Engels S, Cissel DS, Roberts AB, Kondaiah P, Lafyatis R, Sporn MB, Unsicker K (1991) Localization and actions of transforming growth factor-beta s in the embryonic nervous system. *Development* 113: 183–191

47. Hamby ME, Hewett JA, Hewett SJ (2010) Smad3-dependent signaling underlies the TGF-beta1-mediated enhancement in astrocytic iNOS expression. *Glia* 58: 1282–1291

48. Lindholm D, Castren E, Kiefer R, Zafra F, Thoenen H (1992) Transforming growth factor-beta 1 in the rat brain: increase after injury and inhibition of astrocyte proliferation. *J Cell Biol* 117: 395–400

49. Wang X, Chen W, Liu W, Wu J, Shao Y, Zhang X (2009) The role of thrombospondin-1 and transforming growth factor-beta after spinal cord injury in the rat. *J Clin Neurosci* 16: 818–821

50. Klempt ND, Sirimanne E, Gunn AJ, Klempt M, Singh K, Williams C, Gluckman PD (1992) Hypoxia-ischemia induces transforming growth factor beta 1 mRNA in the infant rat brain. *Brain Res Mol Brain Res* 13: 93–101

51. Flanders KC, Ren RF, Lippa CF (1998) Transforming growth factor-betas in neurodegenerative disease. *Prog Neurobiol* 54: 71–85

52. Dobolyi A, Vincze C, Pal G, Lovas G (2012) The neuroprotective functions of transforming growth factor beta proteins. *Int J Mol Sci* 13: 8219–8258

53. Wang G, Yu Y, Sun C, Liu T, Liang T, Zhan L, Lin X, Feng XH (2016) STAT3 selectively interacts with Smad3 to antagonize TGF-beta. *Oncogene* 35: 4388–4398

54. Carow B, Rottenberg ME (2014) SOCS3, a major regulator of infection and inflammation. *Front Immunol* 5: 58

55. Rouillard AD, Gundersen GW, Fernandez NF, Wang Z, Monteiro CD, McDermott MG, Ma'ayan A (2016) The harmonizome: a collection of processed datasets gathered to serve and mine knowledge about genes and proteins. *Database (Oxford)* 2016: baw100

56. Colonna M, Butovsky O (2017) Microglia function in the central nervous system during health and neurodegeneration. *Annu Rev Immunol* 35: 441–468

57. Finak G, McDavid A, Yajima M, Deng J, Gersuk V, Shalek AK, Slichter CK, Miller HW, McElrath MJ, Prlic M *et al* (2015) MAST: a flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. *Genome Biol* 16: 278

58. Huang DW, Sherman BT, Lempicki RA (2008) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* 4: 44

59. Huang DW, Sherman BT, Lempicki RA (2009) Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res* 37: 1–13

60. Hsieh CL, Koike M, Spusta SC, Niemi EC, Yenari M, Nakamura MC, Seaman WE (2009) A role for TREM2 ligands in the phagocytosis of apoptotic neuronal cells by microglia. *J Neurochem* 109: 1144–1156

61. Takahashi K, Rochford CD, Neumann H (2005) Clearance of apoptotic neurons without inflammation by microglial triggering receptor expressed on myeloid cells-2. *J Exp Med* 201: 647–657

62. Zheng H, Jia L, Liu CC, Rong Z, Zhong L, Yang L, Chen XF, Fryer JD, Wang X, Zhang YW *et al* (2017) TREM2 promotes microglial survival by activating Wnt/beta-Catenin pathway. *J Neurosci* 37: 1772–1784

63. Safaiyan S, Kannaiyan N, Snaidero N, Brioschi S, Biber K, Yona S, Edinger AL, Jung S, Rossner MJ, Simons M (2016) Age-related myelin degradation burdens the clearance function of microglia during aging. *Nat Neurosci* 19: 995–998

64. Streit WJ, Braak H, Xue QS, Bechmann I (2009) Dystrophic (senescent) rather than activated microglial cells are associated with tau pathology and likely precede neurodegeneration in Alzheimer's disease. *Acta Neuropathol* 118: 475–485

65. Neves J, Zhu J, Sousa-Victor P, Konjikusic M, Riley R, Chew S, Qi Y, Jasper H, Lamba DA (2016) Immune modulation by MANF promotes tissue repair and regenerative success in the retina. *Science* 353: aaf3646

66. Xia M, Liu J, Wu X, Liu S, Li G, Han C, Song L, Li Z, Wang Q, Wang J *et al* (2013) Histone methyltransferase Ash1 l suppresses interleukin-6 production and inflammatory autoimmune diseases by inducing the ubiquitin-editing enzyme A20. *Immunity* 39: 470–481

67. Kiselev VY, Kirschner K, Schaub MT, Andrews T, Yiu A, Chandra T, Natarajan KN, Reik W, Barahona M, Green AR *et al* (2017) SC3: consensus clustering of single-cell RNA-seq data. *Nat Methods* 14: 483–486

68. Qiu X, Hill A, Packer J, Lin D, Ma YA, Trapnell C (2017) Single-cell mRNA quantification and differential analysis with Census. *Nat Methods* 14: 309–315

69. Losciuto S, Dorban G, Gabel S, Gustin A, Hoenen C, Grandbarbe L, Heuschling P, Heurtaux T (2012) An efficient method to limit

microglia-dependent effects in astroglial cultures. *J Neurosci Methods* 207: 59−71

70. Buttini M, Orth M, Bellosta S, Akeefe H, Pitas RE, Wyss-Coray T, Mucke L, Mahley RW (1999) Expression of human apolipoprotein E3 or E4 in the brains of Apoe-/- mice: isoform-specific effects on neurodegeneration. *J Neurosci* 19: 4867−4880

71. Mazutis L, Gilbert J, Ung WL, Weitz DA, Griffiths AD, Heyman JA (2013) Single-cell analysis and sorting using droplet-based microfluidics. *Nat Protoc* 8: 870−891

72. Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc B Methodol* 57: 289−300

73. Trapnell C, Cacchiarelli D, Grimsby J, Pokharel P, Li S, Morse M, Lennon NJ, Livak KJ, Mikkelsen TS, Rinn JL (2014) The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat Biotechnol* 32: 381−386

Chapter 5

Cellular heterogeneity in a zebrafish model of epilepsy

This chapter includes the manuscript *"New insigths into the early mechanisms of epileptogenesis in a zebrafish model of Dravet syndrome"* published in collaboration with the Esguerra lab (University Oslo, Norway) in Epilepsia [139] where I ,as second author, contributed with the bioinformatics analysis of the sc-RNAseq data and its interpretation. The overarching aim of this study was to identify underlying mechanisms of seizure establishment in Dravet Syndrom (DS) an early-childhood form of epilepsy which is so far not well pharmacological addressable.

5.1 Heterogeneity dynamics during brain development in zebrafish

In this work, new discoveries in identifying and characterizing cellular defects underlying seizure onset during perinatal brain development in zebrafish has given insight into early mechanisms of the Dravet syndrome (DS). The high phenotypic variability among patients makes DS an heterogeneous disease. In this study was introduced a mutation into *scn1lab*, the zebrafish orthologue of *SCN1A*, whose dysfunctions would lead to impairment in interneuron sodium channel activity. The subsequent neuronal hyperexcitability causes seizures. Single-cell RNA-Seq data has widely corroborated and enriched the other complementary analyses such as the electroencephalographic recordings and the pharmacological profiling of mutant larvae. The synaptic activity is supposed to be impaired by an unbalance between excitatory and inhibitory neurotransmitters.

In order to investigate the differences at the gene expression level between the zebrafish $scn1lab^{mut/mut}$

loss-of-function model and zebrafish wild-type (WT), four- and 7 days postfertilization (dpf) brain larvae were pooled and finally subject to brain dissections, cell dissociation, encapsulation of individual cell with beads in accordance with the Drop-Seq protocol and subsequent sequencing. The sc-RNAseq data analysis was carried out by *Monocle* in R [140]. Despite the high presence of noise in the data, based on the t-SNE (T-distributed Stochastic Neighbor Embedding) clustering, I computed a literature based cell type score. The individual score defined for each cell type and quantified by the marker gene expressions allowed to associate each cluster to a specific cell type by the highest percentage of the corresponding score of the cluster (in Fig. **2A,2B** below). This single-cell cell-type inference unveiled the brain composition and quantified relative changes of specific cell populations across condition and time-points. Interestingly, zebrafish *scn1lab*$^{mut/mut}$ larvae displayed dynamic neuronal and glial cell population changes in association with the GABAergic neuronal loss and astrogliosis. In order to investigate the GABAergic and Glutamatergic dynamic profiles and quantify the balance, I defined a score based on gene markers and divided the cells in the neurons cluster into these two categories. The result was a distinct shift in the ration between GABAergic and glutamatergic scores due to the decrease from 4 dpf to 7 dpf to GABAergic mean score in *scn1lab*$^{mut/mut}$ respect to the significant different increase in WT. Whereas a similar trend from 4 dpf to 7 dpf of glutamatergic markers was uncovered between *scn1lab*$^{mut/mut}$ and WT (in Fig. **3A,3B,3C** below). Moreover, the increased number in reactive astrocytes in *scn1lab* mutant relative to WT fish, as resulted from a gliosis score assignment, suggested the hypertrophy of astrocytes as a consequence of epileptic events. Finally, the branching score for the neuronal subpopulation was in agreement with the reduction in dendritic branching of GABAergic neurons prior to seizure onset postulating that structural deficits are established well before the epileptic phenotype arises (in Fig. **3D** below). Overall, this paper provide a first in depth analysis of epileptogenesis in a medical relevant condition.

5.2 Paper related to heterogeneity dynamics in zebrafish brain composition

FULL-LENGTH ORIGINAL RESEARCH

Epilepsia

New insights into the early mechanisms of epileptogenesis in a zebrafish model of Dravet syndrome

Ettore Tiraboschi[1] | Silvia Martina[2] | Wietske van der Ent[1] | Kamil Grzyb[2] |
Kinga Gawel[1,3] | Maria Lorena Cordero-Maldonado[2] | Suresh Kumar Poovathingal[2] |
Sarah Heintz[1] | Somisetty Venkata Satheesh[4] | Jarle Brattespe[5] | Ju Xu[6] |
Maximiliano Suster[7] | Alexander Skupin[2] | Camila V. Esguerra[1,8] [ID]

[1]Chemical Neuroscience Group, Center for Molecular Medicine Norway, University of Oslo, Oslo, Norway

[2]Integrative Cell Signaling Group, Luxembourg Center for Systems Biomedicine, University of Luxembourg, Esch-sur-Alzette, Luxembourg

[3]Department of Experimental and Clinical Pharmacology, Medical University of Lublin, Lublin, Poland

[4]Molecular Toxicology Group, Institute of Basic Medical Sciences, University of Oslo, Oslo, Norway

[5]Department of Biological Sciences, University of Bergen, Bergen, Norway

[6]Department of Biomedicine, University of Bergen, Bergen, Norway

[7]allmyhomes, Berlin, Germany

[8]Department of Pharmacy, University of Oslo, Oslo, Norway

Correspondence
Camila V. Esguerra, Chemical Neuroscience Group, Center for Molecular Medicine Norway, University of Oslo, Gaustadalléen 21, Forskningsparken 0349 Oslo, Norway.
Email: c.v.esguerra@ncmm.uio.no

Present address
Suresh Kumar Poovathingal, Single Cell Analytics & Microfluidics Core, Vlaams Instituut voor Biotechnologie (VIB)-Katholieke Universiteit (KU) Leuven

Abstract

Objective: To pinpoint the earliest cellular defects underlying seizure onset (epileptogenic period) during perinatal brain development in a new zebrafish model of Dravet syndrome (DS) and to investigate potential disease-modifying activity of the $5HT_2$ receptor agonist fenfluramine.

Methods: We used CRISPR/Cas9 mutagenesis to introduce a missense mutation, designed to perturb ion transport function in all channel isoforms, into *scn1lab*, the zebrafish orthologue of *SCN1A* (encoding voltage-gated sodium channel alpha subunit 1). We performed behavioral analysis and electroencephalographic recordings to measure convulsions and epileptiform discharges, followed by single-cell RNA-Seq, morphometric analysis of transgenic reporter-labeled γ-aminobutyric acidergic (GABAergic) neurons, and pharmacological profiling of mutant larvae.

Results: Homozygous mutant (*scn1lab[mut/mut]*) larvae displayed spontaneous seizures with interictal, preictal, and ictal discharges (mean = 7.5 per 20-minute recording; $P < .0001$; one-way analysis of variance). Drop-Seq analysis revealed a 2:1 shift in the ratio of glutamatergic to GABAergic neurons in *scn1lab[mut/mut]* larval brains versus wild type (WT), with dynamic changes in neuronal, glial, and progenitor cell populations. To explore disease pathophysiology further, we quantified dendritic arborization in GABAergic neurons and observed a 40% reduction in arbor number compared to WT ($P < .001$; n = 15 mutant, n = 16 WT). We postulate that the significant reduction in inhibitory arbors causes an inhibitory to excitatory neurotransmitter imbalance that contributes to seizures and enhanced electrical brain activity in *scn1lab[mut/mut]* larvae (high-frequency range), with subsequent GABAergic neuronal loss and astrogliosis. Chronic fenfluramine administration completely restored dendritic arbor numbers to normal in *scn1lab[mut/mut]* larvae, whereas similar treatment with the benzodiazepine diazepam attenuated seizures, but was ineffective in restoring neuronal

Silvia Martina and Wietske van der Ent contributed equally to this work.

Center for Brain and Disease Research, Leuven, Belgium

cytoarchitecture. BrdU labeling revealed cell overproliferation in *scn1lab*[mut/mut] larval brains that were rescued by fenfluramine but not diazepam.

Significance: Our findings provide novel insights into early mechanisms of DS pathogenesis, describe dynamic cell population changes in the *scn1lab*[mut/mut] brain, and present first-time evidence for potential disease modification by fenfluramine.

KEYWORDS

Dravet syndrome, epileptogenesis, fenfluramine, sodium channel, zebrafish

1 | INTRODUCTION

Dravet syndrome (DS) is a severe, intractable, pediatric epileptic encephalopathy.[1] Approximately 70%-80% of patients carry de novo mutations in *SCN1A*, which encodes the voltage-gated sodium channel α1 subunit ($Na_v1.1$).[2] Gene variants for *SCN2A, SCN8A, GABRA1,* or *STXBP1* have also been implicated in DS-like epileptic encephalopathies, making DS a heterologous disease, with high phenotypic variability among patients.[3,4] Seizure onset is typically provoked by fever, usually within the first year of life.[1] Patients exhibit prolonged, frequent, and diverse types of seizures, from febrile or afebrile, to generalized myoclonic and/or absence seizures[5] that are refractory to current antiseizure drugs.

Human and mouse $Na_v1.1$ is predominantly expressed in parvalbumin-positive γ-aminobutyric acid (GABA) interneurons,[6] in brain structures playing a critical role in seizure generation and spread,[7,8] such as the hippocampus or cortex.[9,10] $Na_v1.1$ haploinsufficiency impairs Na^+ currents and action potential firing of GABAergic interneurons, leading to an elevated excitation/inhibition ratio in forebrain structures.[6] Moreover, functional and structural dentate gyrus deficits in the hippocampal network parallel spontaneous seizure onset. In particular, dentate gyrus granule cells exhibited reduced dendritic arborization and increased spine density.[11]

In addition to stiripentol and cannabidiol, now approved by the US Food and Drug Administration as add-on therapy for DS, fenfluramine (FEN) has emerged as a promising therapeutic candidate.[12,13] FEN, an amphetaminelike drug initially introduced into the market as an appetite suppressant, acts as an serotonin (5-hydroxytryptamine [5-HT]) receptor type 2A, 2B, and 2C agonist and a strong inducer of 5-HT release. Recently, FEN was proposed to exert antiseizure activity through allosteric modulation of the sigma-1 receptor.[14] An open-label clinical trial using a clinically confirmed cohort of DS patients indicated that seven of 10 patients taking FEN were seizure-free for >1 year (mean = 6 years[15]). Results from phase 3 clinical trials (FEN hydrochloride; Fintepla, ZX008) showed 64% seizure reduction in children taking 0.8 mg/kg/d,[15] which corroborates the report by Zhang et al,[16] who demonstrated FEN efficacy in decreasing mean duration and frequency of

---

Key Points

- Zebrafish *scn1lab*[mut/mut] larvae display dynamic neuronal and glial cell population changes, including GABAergic neuronal loss and astrogliosis
- *scn1lab*[mut/mut] larvae display strongly reduced dendritic arborization of GABAergic neurons and cell hyperproliferation
- Chronic fenfluramine treatment rescued the observed arborization and cell hyperproliferation defects, whereas diazepam treatment did not

---

epileptiform discharges in a zebrafish *scn1a* knockdown model. Due to concerns associated with FEN and its potential to induce cardiac valve disease and/or pulmonary hypertension,[17] it was withdrawn from the market in 1997. However, the latest retrospective study indicated that long-term treatment with FEN (27 years) did not increase risk of cardiac valvulopathy or pulmonary hypertension.[18]

Zebrafish are now a recognized model of epilepsy, with genetic and chemically induced models able to recapitulate characteristic features of epileptiform activity.[19,20] Here, we introduced an indel mutation into *scn1lab,* the zebrafish orthologue of *SCN1A*. We undertook a comprehensive analysis of *scn1lab*[mut/mut] larvae to obtain insight into the earliest neurodevelopmental and epileptogenic consequences of *scn1lab* dysfunction. Additionally, due to lack of knowledge about the early chronic effects of FEN usage in DS patients, we aimed to clarify whether and how FEN could counteract neurodevelopmental changes in *scn1lab*[mut/mut] brains.

2 | MATERIALS AND METHODS

2.1 | Zebrafish husbandry

Wild-type (WT) adult zebrafish (*Danio rerio*; AB strain; ZIRC) were maintained at 28.5°C on a 14-h/10-h light/dark

cycle under standard aquaculture conditions, and fertilized eggs were collected via natural spawning. Embryos were raised in embryo medium (E3; 1.5 mmol/L HEPES, pH 7.6, 17.4 mmol/L NaCl, 0.21 mmol/L KCl, 0.12 mmol/L MgSO$_4$, and 0.18 mmol/L Ca[NO$_3$]$_2$), under the same conditions as adults. All experiments were approved by the Norwegian Food Safety Authority experimental animal administration's supervisory and application system (FOTS-18/106800-1).

2.2 | Zebrafish lines

The *scn1lab* (c.439_441delATG;443_445delCGC, p.M147_L149delinsI) line, hereafter *scn1lab^mut/mut^*, was generated using CRISPR/Cas9 mutagenesis. The identified founder carried a 6-nucleotide deletion within exon 4 (gRNA sequence: CTGCGCCTTCATGACGCTCAG, Figure S1A). This deletion produced a double amino acid (aa) deletion (Met, Leu) and a single aa change (Thr->Ile; Figure 1A). The aa sequence changed from FMTLSNP to FISNP and affects the ion transport domain (http://www.ebi.ac.uk/interpro/entry/IPR005821; Figure S1B). Dlx5/6:Gal4-T2A-gfp were generated by coinjecting the iTol2_Dlx5/6:Gal4-T2A-gfp plasmid with transposase mRNA. This plasmid was obtained by cloning the dlx6a-1.4kbdlx5a/dlx6a:GFP sequence into a pminiTol2 vector (Addgene #31829).[21] The sequence was taken from the plasmid pcs-dlx6a-1.4kbdlx5a/dlx6a:GFP (gift, Mark
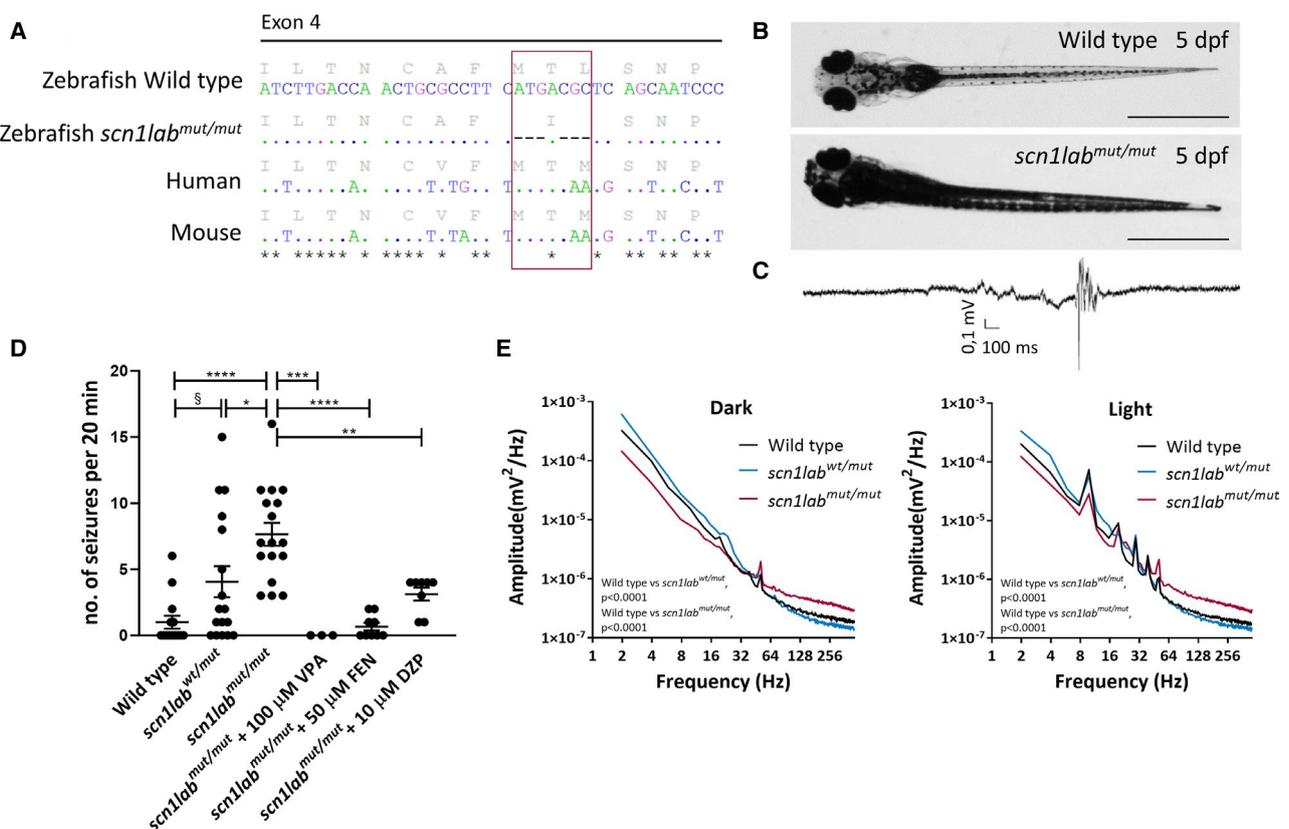
FIGURE 1 Development of *scn1lab^mut/mut^* zebrafish model by CRISPR/Cas9 technique and its characterization. A, Amino acid and nucleotide sequence of *scn1lab* in zebrafish wild type (WT) and *scn1lab^mut/mut^*, and orthologues in human (*SCN1A*) and mouse (*Scn1a*). Dots indicate identical nucleotides, dashes indicate deletions, and stars indicate identical nucleotides in all four sequences. B, WT and *scn1lab^mut/mut^* larvae at 5 days postfertilization (dpf). Increased pigmentation is observed in mutants. Scale bars = 1 mm. C, Example of local field potential (LFP) recording obtained from the *scn1lab^mut/mut^* tectum at 5 dpf, showing a seizure event with preictal and ictal phase. D, *scn1lab^mut/mut^* larvae at 5 dpf show an increased number of seizures over a 20-minute period compared to WT larvae or *scn1lab^wt/mut^* siblings. A low number of seizures were observed in WT larvae, likely caused by needle insertion into the optic tectum. *scn1lab^wt/mut^* displayed a larger number of seizures on average than WT larvae, but more interindividual variability was observed, with some larvae displaying no seizures, whereas others were affected to the same extent as homozygous larvae. No significant difference in the number of seizures was observed between WT larvae and *scn1lab^mut/mut^* larvae treated with valproic acid (VPA), fenfluramine (FEN), or diazepam (DZP), indicating efficacy of these drugs in reducing the number of seizures to background levels. Significance was calculated using one-way analysis of variance with Tukey post hoc test ($*P < .05$, $**P < .01$, $***P < .001$, $****P < .0001$) or unpaired, two-tailed t test ($\S P < .05$). E, Power spectrum analyses of LFP recordings performed under dark conditions, or under a 10-Hz flashing light stimulus. The energy profile of *scn1lab^mut/mut^* larvae differs significantly from WT and *scn1lab^wt/mut^* siblings in both conditions. A Kolmogorov-Smirnov test was used to calculate significance

Ekker). Subsequently, a polymerase chain reaction (PCR) fragment containing Gal4-T2A was inserted in-frame with green fluorescent protein by blunt end ligation into the final plasmid.

2.3 | Locomotor tracking

Locomotor activity was assessed as previously described,[20] with modifications. Four- and 7-days postfertilization (dpf) larvae were placed in 48-well plates (one larva/well) containing 300 µL medium and habituated for 15 minutes to the apparatus (ZebraBox, Viewpoint). Larvae were tracked for 60 minutes in light/dark phase, with 5-minute integration intervals. Measurements were performed at the same daytime period. Distance covered by each larva was recorded in millimeters. Two independent experiments were performed (n = 44-48/group/phase), and the data were pooled.

2.4 | Local field potential recordings

Recordings were obtained from tecta at 4 and 5 dpf as described.[20] Seizure detection was performed through visual inspection and automated using a custom-written R script to minimize bias and artifacts due to muscle contractions. Recorded frequencies were categorized into three bands (1-100, 100-250, and 250-500 Hz). If amplitude exceeded 3× background, the event was considered a seizure. This method was based on high-frequency oscillations (>100 Hz) as a reported marker of epileptic activity.[22] Power spectrum was analyzed using Clampfit 10.2 software (Molecular Devices). Four-minute-long recordings were used for computing the power spectrum from larvae at 5 dpf, and each condition was averaged per group. Light stimuli were generated with a custom-made device comprising an Arduino board sending a 10-Hz sinusoidal wave to a transistor, which in turn powered a 2-W light-emitting diode emitting white light. Light was delivered via optic fiber to the recording stage.

2.5 | Real-time quantitative PCR

Heads of 6-dpf anesthetized larvae were collected in pools of 10 for RNA extraction. RNA was purified using TRIzol, and cDNA was synthesized using the SuperScript IV First-Strand Synthesis System (Invitrogen). cDNA was amplified using PowerUp SYBR Green Master Mix (Applied Biosystems) according to the manufacturer's instructions. Relative enrichment was computed according to the $2^{-\Delta\Delta t}$ method.[23] Expression levels were normalized against β-actin. Primer sequences were actb2_f_5'TTCTCTTAA

GTCGACAACCCCC3', actb2_r_5'ACAATACCGTGCT CGATGGG3'; elavl3_f_5'ATCAACACGCTCAACGGT CT3', elavl3_r_5'TTACCAGGATGCGTGAGGTG3'; gad 67_f_5'TGTGTCCGATGGCTTGAGTC3', gad67_r_5'CA CACGGAGGATGGTTCACA3'; and vglut1_f_5'CGGCT CATTCTTCTGGGGTT3', vglut1_r_5'GACCATGATCA CACACCCGT3'.

2.6 | Single-cell transcriptomics

Brain dissections, cell dissociations, encapsulation, and generation of microfluidic Drop-Seq chips were performed in accordance with the Drop-Seq protocol[24] as described (detailed in Supplementary Methods in Data S1).

2.7 | Bioinformatics

Sequencing data were processed by the Drop-Seq bioinformatics pipeline (v1.13), and reads were mapped against genome version GRCz11.92, resulting in corresponding digital expression matrices. Data analysis was performed by Monocle (v2.4.0) in R (v3.4.4) following the standard workflow for quality control, dimension reduction, and differential expression analysis (see Supplementary Methods in Data S1).

2.8 | Imaging and quantification of arbors

To visualize GABAergic neurons, *scn1lab*^*wt/mut* fish were crossed with Dlx5/6:Gal4-T2A-gfp/UAS:nfsb-mCherry reporter line. Adults were then in-crossed for generating *scn1lab*^*mut/mut*. For controls, Dlx5/6:Gal4-T2A-gfp/ UAS:nfsb-mCherry fish were crossed with WT. Larvae were treated from 1 dpf with 0.003% phenylthiourea, to prevent pigmentation. FEN (50 µmol/L) or diazepam (DZP; 10 µmol/L) was added to E3 medium at 3 dpf and replaced daily until 6 dpf. Larvae were anesthetized in 0.001% tricaine (Sigma), fixed for 3 hours at room temperature with 4% paraformaldehyde, mounted on glass slides, and imaged using confocal microscopy. A dorsal z-stack of the optic tectum was collected using a ×40 lens and a z-resolution of 0.44 µm. For Sholl analysis,[25,26] images were filtered using the 3D-Median filter in ImageJ. A z-projection of the tectum was generated, and the resulting image was converted to a thresholded binary image. The extent of arborization was quantified using Sholl analysis (plug-in; http://imagej.net/Sholl_Analysis). The number of intersections was normalized against the number of neurons quantified in the imaged area, and this value was used for statistical analysis.

2.9 | BrdU immunohistochemistry

BrdU labeling was performed as described in Supplementary Methods (Data S1).

2.10 | Statistical analysis

Data were analyzed using GraphPad Prism 7.05. For multiple comparisons, one-way analysis of variance followed by Tukey or Sidak post hoc test was used. Students unpaired *t* test or Kolmogorov-Smirnov test was performed if necessary. For single-cell RNA-Seq (scRNA-Seq) analysis, Bonferroni compensation for multiple testing was used when applicable.

3 | RESULTS

3.1 | *scn1lab*^*mut/mut* larvae recapitulate features of DS

We generated a zebrafish line, using CRISPR/Cas9 mutagenesis, carrying a double aa deletion and single aa substitution in the first ion transport domain of *scn1lab* (Figure S1). Morphologically, *scn1lab*^*mut/mut* larvae phenocopied previously described zebrafish *scn1lab* loss-of-function models, displaying hyperpigmentation from 4 dpf and uninflated swim bladders.[16,27] At 5 dpf, a "kink" posterior to the trunk became evident and larvae struggled to maintain an upright posture (Figure 1B). Homozygotes survive until 14 dpf, whereas heterozygotes breed well and remain healthy up to 18 months. We assessed locomotor activity[16,19,20] at 4 and 7 dpf. *scn1lab*^*mut/mut* larvae were more active and traveled greater distances in light phase compared to WT. However, whereas WT siblings tested in dark phase increased locomotor activity, *scn1lab*^*mut/mut* larvae displayed minimal increase in movement in dark conditions (Figure S2). Touch response in 5-dpf *scn1lab*^*mut/mut* larvae was comparable to WT, with delayed touch response in two of 24 mutant larvae versus one of 24 WT larvae (data not shown). Tectal field recordings indicated that both *scn1lab*^*mut/mut* and *scn1lab*^*wt/mut* larvae displayed seizures starting from 4 dpf (Figure 1D, Figure S3). Five-days-postfertilization *scn1lab*^*mut/mut* larvae (n = 17) displayed spontaneous electrographic discharges with high amplitude and duration (Figure 1C,D), which were inhibited by acute administration of 100 μmol/L valproic acid (VPA; $P < .001$; n = 3), 50 μmol/L FEN ($P < .0001$; n = 9), or 10 μmol/L DZP ($P < .01$; n = 8; Figure 1D). We then performed electroencephalography (EEG) with alternating periods of stimulus deprivation or photostimulation, by recording larvae in the dark or delivering a 10-Hz sine wave flashing light. Spectral EEG analyses revealed that *scn1lab*^*mut/mut* larval (n = 11) energy profiles differed significantly from heterozygous (n = 8) and WT counterparts ($P < .0001$; n = 11).
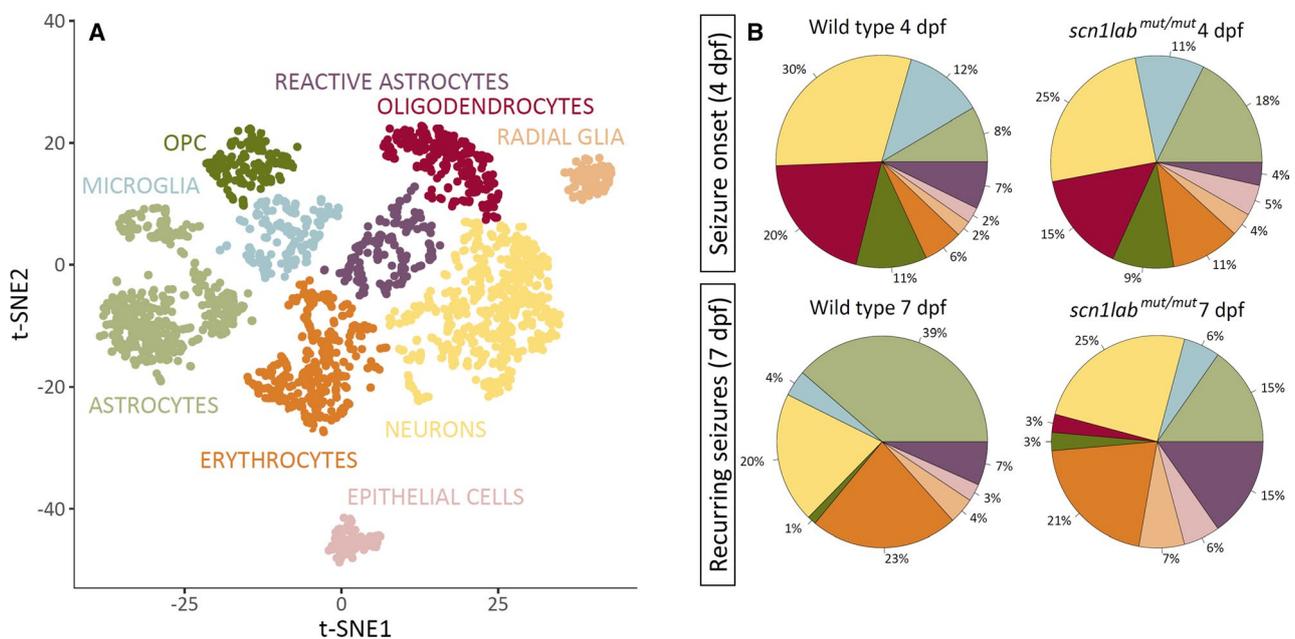
FIGURE 2 Single-cell RNA-Seq data reveals changes in brain composition. A, The 2508 clearly identified cells were clustered and cell identities inferred by cell type-specific scores (Figure S4) based on marker genes (Table S1). Interestingly, distinct astrocytic clusters were identified that indicate maturation of cells (Figure S4). OPC, oligodendrocyte progenitor cells. B, The brain composition exhibits an increase in radial glia at 7 days postfertilization (dpf) for both cell lines and a doubled percentage of reactive astrocytes in *scn1lab*^*mut/mut* cells. The strongly reduced number of oligodendrocytes at 7 dpf probably results from damage during cell isolation and corresponding leakage of mRNAs (Figure S5). Cell clusters presented here are from pooled larval brains derived from wild type or homozygotes. tSNE, T-distributed Stochastic Neighbor Embedding

In particular, energy distribution at high frequencies (>100 Hz) was greater (Figure 1E).

3.2 | Brain composition analysis

We performed scRNA-Seq analysis using our Drop-Seq pipeline (see Supplementary Methods in Data S1). Epithelial cells formed a distinct subpopulation (light red cluster), and neuronal- and glia-related clusters were separated well. Astrocytes formed subclusters, and radial glia exhibited a specialized expression profile leading to a distinct cluster (Figure 2A). Comparison of cell type matching with sample identity (Figure S5) did not indicate a *scn1lab*[mut/mut]-specific neuronal expression profile but reduced astrocytic maturation. Increased vascularization became apparent from the increased number of erythrocytes at 7 dpf (Figure 2B). Oligodendrocyte absence at 7 dpf for both conditions is related to the exclusion of leaking cells, as the majority of these were associated with 7-dpf samples and careful evaluation of the corresponding expression profiles indicated their oligodendrocytic origin by the remaining her4.1 transcripts (Figure S5). The relative decrease in the neuronal subpopulation was also caused by an increase in astrocytes in both lines where, in particular, radial glia appeared to be enriched in *scn1lab*[mut/mut] larvae.

3.3 | Reduced GABAergic neurons and increased gliosis

We investigated GABAergic and glutamatergic neurons by calculating subtype-specific scores based on relative marker gene expression (Table S1, Figure S6). Glutamatergic scores were similar for *scn1lab*[mut/mut] and WT, exhibiting an increase from 4 to 7 dpf, where only the increase
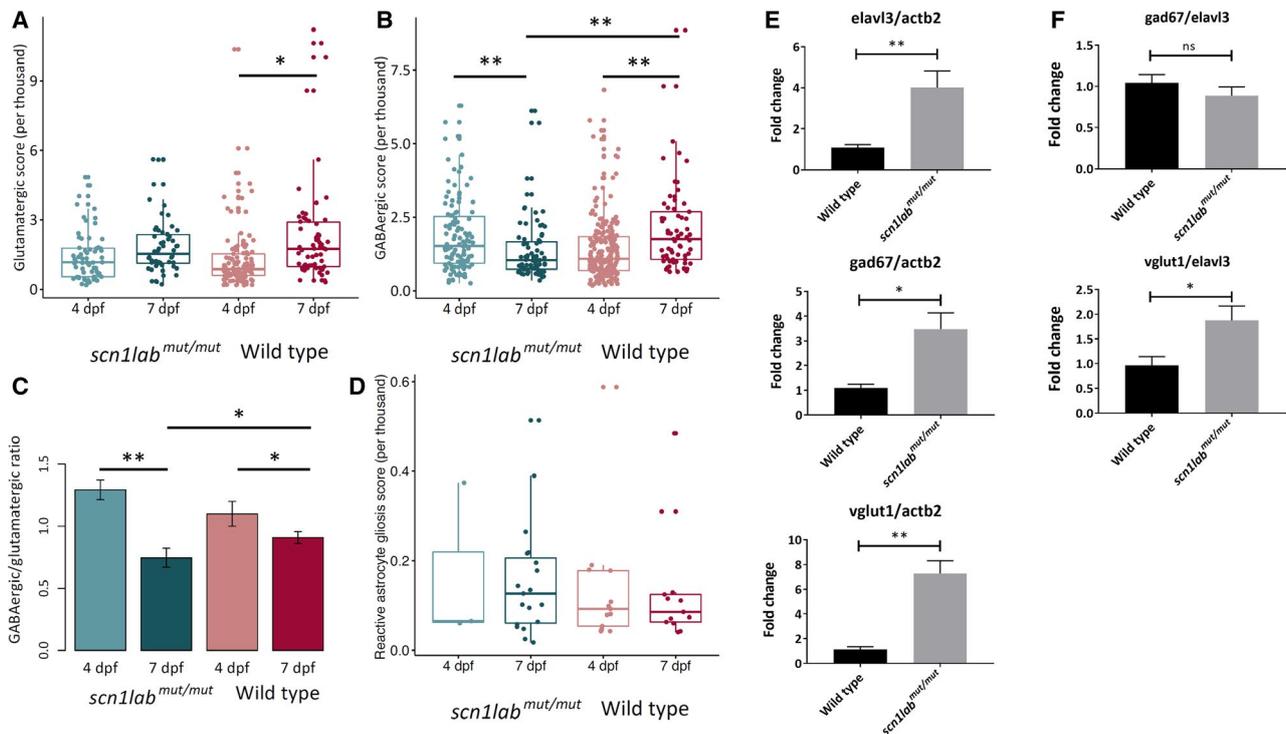
FIGURE 3 Single-cell RNA-Seq data reveal changes in γ-aminobutyric acidergic (GABA) and glutamatergic profiles defined by corresponding marker genes during development and between conditions. A, The glutamatergic score indicates a stronger increase of glutamatergic profiles in wild-type (WT) than in *scn1lab*[mut/mut] larvae from 4 to 7 days postfertilization (dpf). B, The GABA-related score indicates a significant decrease in neurons of *scn1lab*[mut/mut] larvae and an increase in WT from 4 to 7 dpf, leading to a significant difference between the conditions. C, Combining these scores to investigate the balance between GABA- and glutamate-related neurons shows a dramatic change in *scn1lab*[mut/mut] larvae, where the ratio between GABAergic and glutamatergic profiles is reduced by nearly 50% at 7 dpf compared to 4 dpf, whereas in WT only a minor reduction is observed. This results in a significant difference between *scn1lab*[mut/mut] and WT larvae at 7 dpf and may indicate an underlying mechanism of epileptogenesis. D, The increased number of reactive astrocytes found in *scn1lab*[mut/mut] do also exhibit an increased gliosis activity defined by corresponding marker genes (Table S1) at 7 dpf, as indicated by differentially expressed gene analysis. E, At 6 dpf, *scn1lab*[mut/mut] larvae had an increased expression of glutamatergic (*vglut1*), GABAergic (*gad67*), and postmitotic (*elavl3*) neuronal markers compared to WT siblings, as measured by real-time quantitative polymerase chain reaction. F, Upregulation of the excitatory system (as observed by normalization of *vglut1* over *elavl3*), but not the inhibitory system (as observed by normalization of *gad67* over *elavl3*), indicates a shift in the GABA/GLUT population ratios in the brain of *scn1lab*[mut/mut] larvae. $*P < .05$, $**P < .01$. ns, not significant
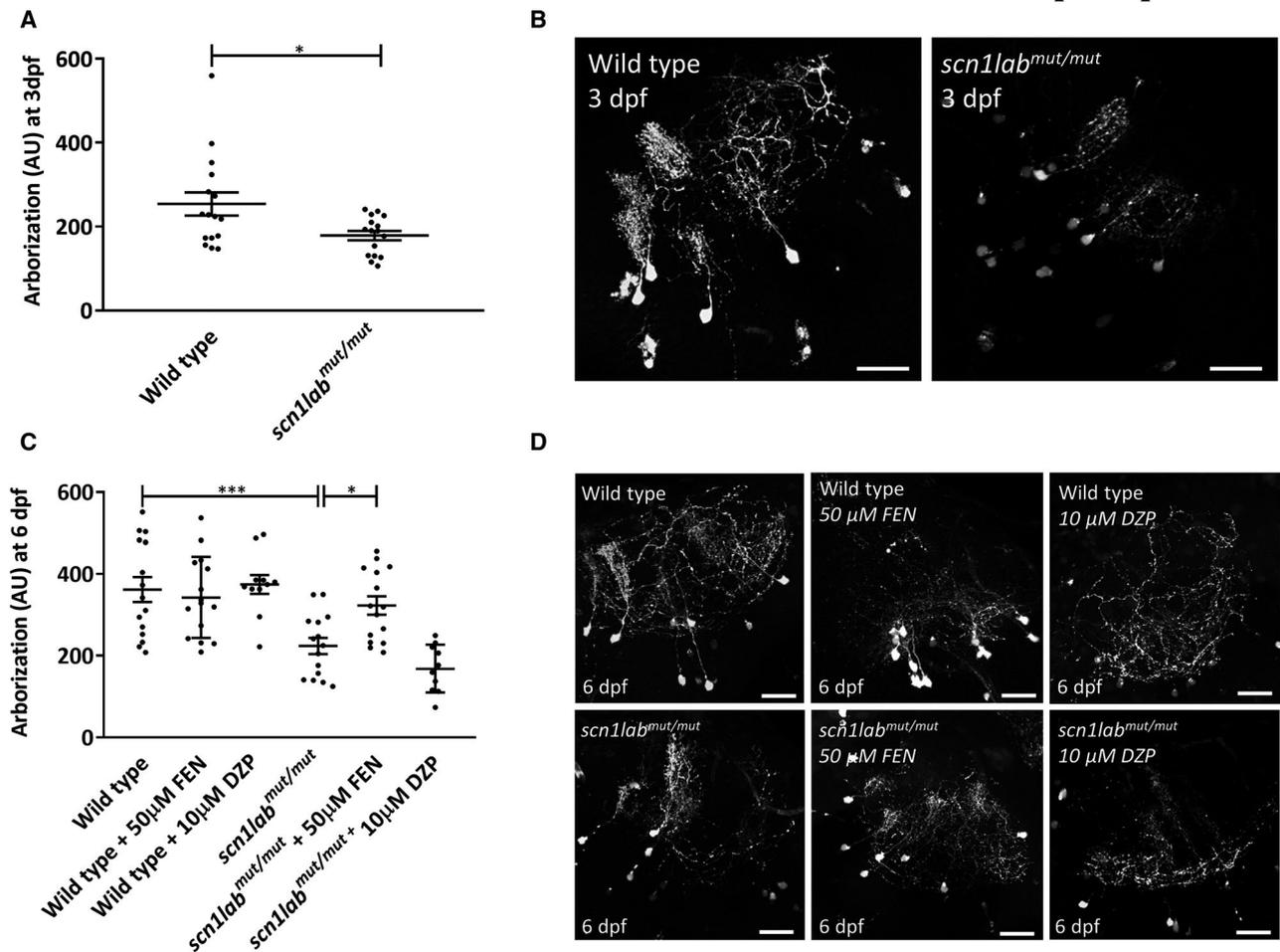
FIGURE 4 Dendritic arborization in *scn1lab*^mut/mut larvae. A, *scn1lab*^mut/mut larvae show a reduced number of dendritic arbors at 3 days postfertilization (dpf). Significance of differences was calculated using an unpaired, two-tailed *t* test. *$P < .05$. B, Representative images of dendritic arbors in the neuropil of wild-type (WT) and *scn1lab*^mut/mut larvae at 3 dpf. Scale bars = 25 µm. C, Reduction in dendritic arborization in *scn1lab*^mut/mut larvae observed at 6 dpf can be rescued by the addition of 50 µmol/L of fenfluramine (FEN), but not 10 µmol/L of diazepam (DZP), from 3 dpf until 6 dpf. Significance of differences was calculated using one-way analysis of variance with Sidak post hoc test. *$P < .05$, ***$P < .001$. D, Representative images of dendritic arbors in the neuropil of WT and *scn1lab*^mut/mut larvae at 6 dpf, with or without prior incubation in 50 µmol/L of FEN or 10 µM of DZP from 3 dpf onward. Scale bars = 25 µm

in WT was statistically significant (Figure 3A). In contrast, GABAergic scores revealed significant differences during development and between WT and mutant. Whereas the GABAergic score of *scn1lab*^mut/mut larvae showed a significant reduction from 4 to 7 dpf, WT larvae exhibited a significant increase in GABAergic marker levels at 7 dpf (Figure 3B). Consequently, *scn1lab*^mut/mut larvae had significantly reduced GABAergic marker expression at 7 dpf compared to WT. The similar expression of glutamatergic markers and the significant differences in the GABAergic score led to a distinct shift in the ratio between GABAergic and glutamatergic scores (Figure 3C). Whereas the ratio was slightly reduced in WT (10%), *scn1lab*^mut/mut larvae exhibited a reduction of 43% at 7 dpf and a significant decrease (19%) in WT at 7 dpf.

To investigate the increased number of reactive astrocytes in *scn1lab*^mut/mut larvae at 7 dpf further, we performed differentially expressed gene analysis and found an enrichment of gliosis-related genes. We calculated a gliosis score based on marker genes (Table S1) and found a trend for higher gliosis activity in reactive astrocytes of *scn1lab*^mut/mut larvae (Figure 3D, Figure S6). To test whether the decrease in GABAergic cells correlated with a change in neuronal morphology at the expression level, we targeted axonogenesis-related genes (Table S1) in the scRNA-Seq data. Although we could not find a significant effect on branching expression levels, we observed a trend toward downregulation of neuronal branching genes in the mutant (18%), whereas WT larvae exhibited a nonsignificant reduction of 6% (Figure S7).

To confirm that the distinct EEG energy profile of *scn1lab*^mut/mut is caused by increased synaptic activity due to reduced inhibitory tone, we used real-time quantitative PCR to quantify glutamatergic (*vglut1*), GABAergic

(*gad67*), and postmitotic (*elavl3*) neuronal markers in larval brains. Homozygous *scn1lab*^mut/mut had increased expression of all markers compared to WT ($P < .001$, *elavl3/actb2* and *vglut1/actb2*; $P < .05$, *gad67/actb2*; n = 7; Figure 3E). However, normalization of *vglut1* expression against *elavl3* indicated upregulation of excitatory markers ($P < .05$), whereas normalization of *gad67* expression against *elavl3* showed no differences between *scn1lab*^mut/mut and WT ($P > .05$; Figure 3F).

3.4 | Neuronal defects restored by FEN

The amplitude difference in the power spectrum observed between *scn1lab*^mut/mut and WT indicated possible alteration in optic tectum cytoarchitecture. We therefore quantified dendritic arborization in GABAergic neurons, as reduced branching would explain reduced inhibitory tone. Sholl analysis revealed a 30% decrease ($P < .05$; n = 16) in dendritic arborization of GABAergic tectal neurons of 3 dpf *scn1lab*^mut/mut larvae compared to WT (n = 16; Figure 4A,B). By 6 dpf, dendritic arbors were reduced by 40% ($P < .001$; n = 15; Figure 4C,D). As acute FEN treatment effectively suppressed seizures in previously described zebrafish DS models,[16,28] we tested whether chronic FEN administration might affect dendritic morphology. We administered 50 μM FEN chronically for 3 consecutive days, starting at 3 dpf,

prior to seizure onset. Chronic treatment of *scn1lab*^mut/mut larvae with FEN significantly ($P < .05$; n = 15) abrogated the reduction in dendritic arborization, whereas chronic treatment of WT did not induce any changes ($P > .05$; n = 15; Figure 4C,D). Chronic treatment of *scn1lab*^mut/mut larvae with DZP attenuated seizures but showed no amelioration of reduced arborization (Figure 4C,D).

We also evaluated changes in cell proliferation through BrdU labeling in *scn1lab*^mut/mut brains. We observed an increase ($P < .01$; n = 8) in BrdU-positive cells, compared to WT at 6 dpf (n = 16; Figure 5). To test whether FEN might also regulate cell proliferation, we exposed *scn1lab*^mut/mut larvae to FEN as described above. FEN restored ($P < .05$; n = 9) the number of BrdU-positive tectal cells to untreated WT baseline. FEN treatment of WT also increased ($P < .0001$; n = 7) BrdU-positive cells when compared to untreated WT controls (Figure 5).

4 | DISCUSSION

4.1 | New zebrafish mutant to study DS epileptogenesis

The *scn1lab* mutant described here replicates classic features of DS and recapitulates both previously described phenotypes for other zebrafish *scn1lab* loss-of-function models and *Scn1a* mouse mutants.[4,6,27,28] *didy* homozygotes, however,
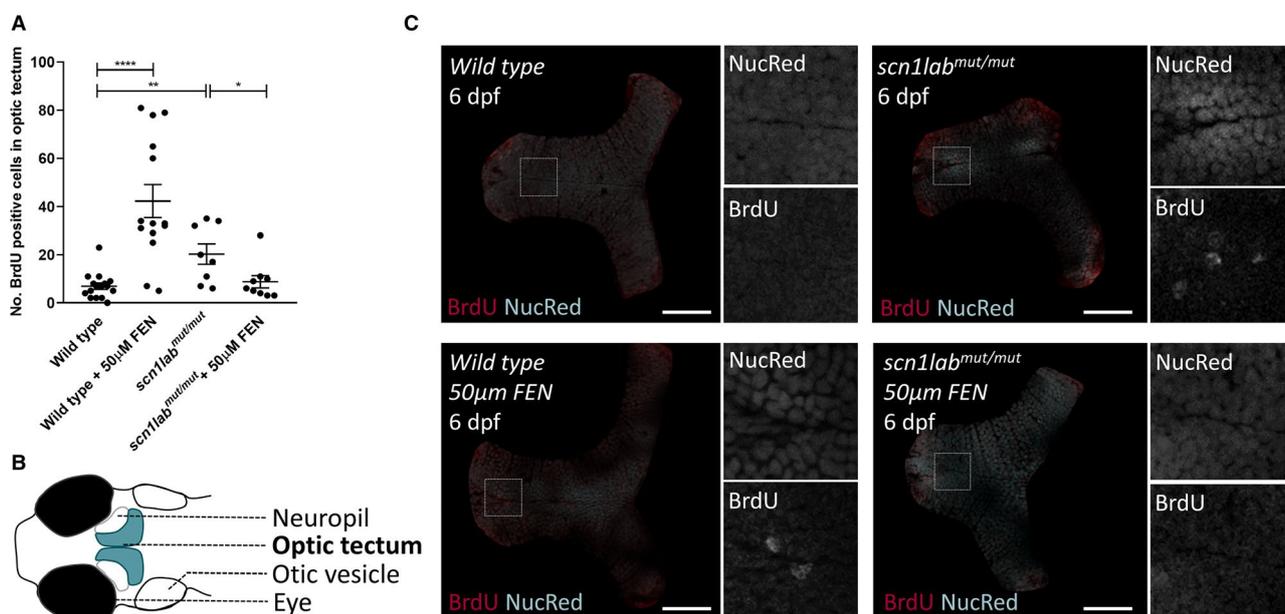
FIGURE 5 Proliferation of cells in the optic tectum of *scn1lab*^mut/mut larvae. A, At 6 days postfertilization (dpf), *scn1lab*^mut/mut larvae showed an increased number of proliferating cells in the optic tectum, as measured by BrdU-positive cells in 30-μm-thick optical sections. Incubation of larvae in 50 μmol/L of fenfluramine (FEN) from 3 dpf onward resulted in an increased number of BrdU-positive cells in wild-type (WT) larvae at 6 dpf, while reducing the number to untreated WT levels in the *scn1lab*^mut/mut larvae. Statistical significance was calculated using multiple unpaired, two-tailed *t* tests. *$P < .05$, **$P < .01$, ****$P < .0001$. B, Cartoon representation of the head of a 6-dpf larva, indicating the position of the optic tectum where proliferation was measured. C, Representative z-slices of masked optic tecta with BrdU staining of WT and *scn1lab*^mut/mut larvae at 6 dpf, with or without prior incubation in 50 μmol/L of FEN from 3 dpf onward. Squares indicate the location of the zoomed-in, split-channel images to the right. Scale bars = 50 μm

survive until 10 dpf,[28] whereas *scn1lab^mut/mut* larvae survive until 14 dpf. Moreover, seizures were only reported in *didy* homozygotes, whereas seizures were also observed in *scn-1lab^wt/mut* heterozygotes, possibly attributable to differences in genetic background or the nature of the mutation (nonsense vs missense).

4.2 | Increased brain excitability

The higher energy distribution observed in *scn1lab^mut/mut* brains, both after photostimulus and under dark conditions, indicated that the summation signal of dendritic potentials was excitatory. This is not surprising, given the development of spontaneous seizures, but the finding that this is mostly represented in the high-frequency gamma band was suggestive not only of increased synaptic activity but possibly of decreased inhibitory tone. Neuronal hyperactivity, driven primarily by reduced background inhibition, can be caused by various mechanisms. For example, abnormal gamma-band activity has been observed in schizophrenia patients and related animal disease models as a result of impaired GABAergic neurotransmission, particularly in parvalbumin-expressing neurons.[29] Regarding *SCN1A*-derived disorders, two mechanisms underlying neuronal hyperexcitability have been proposed: impaired interneuron sodium channel activity or increased pyramidal neuron excitability.[30] With regard to the former, studies on mouse *Scn1a* mutants support the hypothesis of disinhibition resulting from decreased $Na_v1.1$ function in interneurons.[9,31]

4.3 | GABAergic neuron loss

There is sufficient evidence that loss of GABAergic neurons is a consistent finding among all models of acquired epilepsy including kindling, status epilepticus, and traumatic brain injury as well as a frequently observed hallmark of temporal lobe epilepsy (TLE).[32–37] Furthermore, selective loss of GABAergic neurons can also give rise to an epileptic state in genetic models. Mice with loss of *Dlx1*, which encodes a transcription factor regulating GABAergic interneuron development, develop epilepsy as a result of a time-dependent reduction in the number of these neurons within the cerebral cortex and hippocampus.[38]

A previous study reported no evident loss of GABAergic neurons in the zebrafish *scn1lab* mutant *didy*.[27] Possible explanations for this discrepancy in cellular phenotype between the models are differences in mutation type and/or the genetic background. Another plausible explanation is that we analyzed GABAergic neuron population dynamics at a later developmental stage (7 dpf instead of 5 dpf as carried out in the *didy* study) and therefore after a longer period of spontaneous

recurring seizures. Our data at 4 dpf corroborate with the *didy* study findings of no evident cell loss. Interestingly, our data indicate that there is a slight increase in GABAergic neurons at this earlier time point in the mutant, indicating changes in cell proliferation in the mutant during brain development. Furthermore, we found no difference in glutamatergic marker levels between the conditions but an increase from 4 to 7 dpf. Thus, by 7 dpf, the decrease in the GABAergic neuron signature and the increase in glutamatergic signature indicated a shift in the GABA/GLUT ratio in *scn1lab^mut/mut* larvae, providing additional evidence for decreased inhibitory tone.

Importantly, a proportion of GABAergic neurons still remain in mutant larvae, as has been observed in *Scn1a* mouse models, other models of acquired epilepsy, and human TLE. Alterations in the remaining GABAergic neurons of these mammalian models have been proposed as contributors to epileptogenesis. Critical changes resulting in the establishment of an epileptic state include impaired function of remaining GABAergic neurons and morphological changes that eventually lead to aberrant circuitry.[39]

4.4 | Reduced dendritic arborization

The significant reduction in dendritic branching of GABAergic neurons prior to seizure onset in our DS model shows that structural deficits are established well before the epileptic phenotype arises. This indicates that seizures per se are not the primary cause of the observed arborization defects and that the converse may be true; reduced arborization is a possible primary mechanism underlying seizures in DS. Importantly, these defects arose when GABA was still excitatory, that is, prior to the GABA developmental switch, which occurs at 2.5 dpf in zebrafish. The early branching defects in GABAergic neurons are thus likely to be due to the inability of these neurons to propagate action potentials, which in turn impairs strengthening and pruning of synaptic connections. In general, dysfunctional voltage-gated sodium channels would lead to impaired action potentials and current in GABAergic inhibitory neurons,[6,39] lowering inhibitory tone and consequently increasing overall network excitability, which would dramatically alter plasticity, morphology, network dynamics, and functional topology of the neural circuitry. A similar mechanism was proposed for the zebrafish *scn1lab* mutant *didy*, originally described as displaying an abnormal optokinetic response due to its inability to sustain saccade eye movements when presented with a repeating visual stimulus pattern.[40] The authors postulated that depleting *scn1a* channel function in *didy* could potentially reduce channel density, in turn resulting in prolonged stimulation that would eventually deplete the pool of activatable channels and ultimately render neurons unexcitable.

4.5 | FEN treatment rescues dendritic arborization defects

FEN, a serotonin 5-HT$_{2A}$ and 5-HT$_{2C}$ receptors agonist, is highly effective in diminishing seizures in reported zebrafish DS models. Surprisingly, there are no reports describing effects of chronic or acute FEN treatment in DS rodent models. A recent study that evaluated selective 5-HT$_{2C}$ agonists lorcaserin, CP809101, and FEN in multiple acute rodent seizure models (maximum electroshock seizure [MES], MES threshold, 6-Hz electrical convulsive seizure, pentylenetetrazole, and amygdala kindling), showed that with the exception of FEN-mediated tonic seizure inhibition in the MES rat test, there was no antiseizure effect for all compounds in all other tests.[41] The authors concluded that early developmental genetic models might be more appropriate for study of certain drug leads, indicating a need to consider integrating additional animal models into preclinical drug discovery pipelines.

Chronic treatment with FEN prevents cell overproliferation in *scn1lab*$^{mut/mut}$ larvae. However, in WT larvae it increases cell proliferation. These results corroborate previous findings that showed neurogenesis-stimulating activity of antidepressants in the dentate gyrus.[42] The role of serotonin in regulating interneuron morphology has previously been shown. Inhibition of serotonin synthesis in rat embryos with para-chlorophenylalanine disrupted pyramidal neuron maturation within the somatosensory cortex by reducing dendritic arborization.[43] Benzodiazepines such as DZP and clobazam have been used to treat DS.[18] Our data suggest that FEN activity is specific and not simply due to seizure inhibition, as chronic treatment using DZP did not restore dendritic arbor numbers to normal.

Dendritic arborization defects have also been reported for *Scn1a*$^{E1099X/+}$ and *Scn1b*$^{C121W/+}$ mice.[11,44] *Scn1a*$^{E1099X/+}$ mice display progressively reduced dendritic arborization and excessive spines in GABAergic hippocampal dentate gyrus neurons, whereas *Scn1b*$^{C121W/+}$ mice exhibit reduced arborization in subicular pyramidal neurons. These morphological abnormalities, however, were analyzed coincident with (ie, not prior to) developmental spontaneous seizure onset. It is therefore unclear whether arborization defects in these models were a key trigger for epileptogenesis or a consequence thereof. Clearly, further investigation is warranted to determine the exact role of dendritic arborization in DS pathogenesis. Such studies would require dendrite morphology analysis in our DS model at later developmental stages, to allow for sufficient drug washout after chronic FEN or DZP administration (to rule out acute seizure inhibition from residual drug), ideally resulting in significant seizure reduction or freedom in FEN-rescued but not DZP-treated larvae.

4.6 | Gliosis

Radial glial cells are more abundant in *scn1lab* mutants compared to WT, with a doubling of reactive astrocytes in the mutant relative to WT, indicating a trend toward increased gliosis. Reactive astrocytes increase in number as a consequence of epileptic events but are also suggested to play a functional role in the causation of seizures, for instance, after brain trauma.[45] Our scRNA-Seq results are in line with a previous study that compared RNA-Seq data obtained from *Scn1a*$^{\pm}$ mice on two different strains. Half of differentially expressed genes shared between the two models were associated with astrogliosis, including genes upregulated in response to human and rodent seizures or traumatic brain injury.[46]

As mentioned, heterozygous *scn1lab* mutant larvae described in this study also display frequent seizures yet are fertile and survive to adulthood. This could prove useful for analyzing *Scn1a* dysfunction during later stages of epileptogenesis for studying comorbid features and for testing disease-modifying activity of drugs in the context of a mature brain. The astrogliosis phenotype is also worth characterizing further. Notably, a study that performed electrophysiological recordings on acute brain slices from *Scn1a*$^{\pm}$ mice found that, at later developmental stages, parvalbumin-positive fast-spiking basket cell interneuron excitability normalized. The results of this study implied that impairing action potential generation by parvalbumin-positive fast-spiking basket cell interneurons contribute to the initial epilepsy phenotype but may not be the sole mechanism that drives later, chronic epilepsy in DS.[47] It is equally important, however, to consider that the variable severity and expressivity of the murine *Scn1a*$^{\pm}$ epilepsy phenotype is highly strain dependent.[48,49] Therefore, it will be interesting to explore whether stabilization of interneuron activity can be recapitulated in the zebrafish *scn1lab*$^{wt/mut}$.

This study describes a new zebrafish *scn1a* channelopathy model that recapitulates many aspects of DS and may prove useful in identifying other disease mechanisms such as the role of neuron-glia interaction and glutamatergic neurons in the process of epileptogenesis or the development of other associated comorbidities. These data illustrate the utility of zebrafish genetic epilepsy models in elucidating early mechanisms leading to an epileptogenic state. The remarkable efficacy of FEN in this DS model demonstrates the utility of zebrafish in uncovering potential disease-modifying activity of drug leads.

CONFLICT OF INTEREST

None of the authors has any conflict of interest to disclose. We confirm that we have read the Journal's position on issues involved in ethical publication and affirm that this report is consistent with those guidelines.

DATA AVAILABILITY STATEMENT

Supporting data are available as supplementary materials or, upon request, from the corresponding author. scRNA-Seq data will be made available online through Gene Expression Omnibus.

ORCID

Camila V. Esguerra https://orcid.org/0000-0002-2271-8094

REFERENCES

1. Dravet C. The core Dravet syndrome phenotype. Epilepsia. 2011;52(2):3–9.
2. Claes L, Del-Favero J, Ceulemans B, et al. De novo mutations in the sodium-channel gene SCN1A cause severe myoclonic epilepsy of infancy. Am J Hum Genet. 2001;68:1327–32.
3. Steel D, Symonds JD, Zuberi SM, Brunklaus A. Dravet syndrome and its mimics: beyond SCN1A. Epilepsia. 2017;58:1807–16.
4. Marini C, Scheffer IE, Nabbout R, et al. The genetics of Dravet syndrome. Epilepsia. 2011;52(2):24–9.
5. Harkin LA, McMahon JM, Iona X, et al. The spectrum of SCN1A-related infantile epileptic encephalopathies. Brain J Neurol. 2007;130:843–52.
6. Ogiwara I, Miyamoto H, Morita N, et al. Nav1.1 localizes to axons of parvalbumin-positive inhibitory interneurons: a circuit basis for epileptic seizures in mice carrying an Scn1a gene mutation. J Neurosci. 2007;27:5903–14.
7. Rutecki PA, Grossman RG, Armstrong D, Irish-Loewen S. Electrophysiological connections between the hippocampus and entorhinal cortex in patients with complex partial seizures. J Neurosurg. 1989;70:667–75.
8. Chabardès S, Kahane P, Minotti L, et al. The temporopolar cortex plays a pivotal role in temporal lobe seizures. Brain J Neurol. 2005;128:1818–31.
9. Kelsom C, Lu W. Development and specification of GABAergic cortical interneurons. Cell Biosci. 2013;3:19.
10. Pelkey KA, Chittajallu R, Craig MT, Tricoire L, Wester JC, McBain CJ. Hippocampal GABAergic inhibitory interneurons. Physiol Rev. 2017;97:1619–47.
11. Tsai MS, Lee ML, Chang CY, et al. Functional and structural deficits of the dentate gyrus network coincide with emerging spontaneous seizures in an Scn1a mutant Dravet syndrome model during development. Neurobiol Dis. 2015;77:35–48.
12. Ceulemans B, Boel M, Leyssens K, et al. Successful use of fenfluramine as an add-on treatment for Dravet syndrome. Epilepsia. 2012;53:1131–9.
13. Polster T. Individualized treatment approaches: fenfluramine, a novel antiepileptic medication for the treatment of seizures in Dravet syndrome. Epilepsy Behav. 2019;91:99–102.
14. Martin P, Maurice T, Gammaitoni A, et al. Fenfluramine has in vivo activity as a positive allosteric modulator of sigma-1 receptors. Poster presented at: American Epilepsy Society Annual Meeting; December 1-5, 2017; Washington, DC.
15. Lagae L, Sullivan J, Cross H, et al.ZX008 (fenfluramine) in Dravet syndrome: results of a phase 3, randomized, double-blind, placebo-controlled trial. Poster presented at: American Epilepsy Society Annual Meeting; December 1-5, 2017; Washington, DC.
16. Zhang Y, Kecskés A, Copmans D, et al. Pharmacological characterization of an antisense knockdown zebrafish model of Dravet syndrome: inhibition of epileptic seizures by the serotonin agonist fenfluramine. PLoS One. 2015;10:e0125898.
17. Connolly HM, Crary JL, McGoon MD, et al. Valvular heart disease associated with fenfluramine-phentermine. N Engl J Med. 1997;337:581–8.
18. Knupp KG, Wirrell EC. Treatment strategies for Dravet syndrome. CNS Drugs. 2018;32:335–50.
19. Baraban SC, Taylor MR, Castro PA, Baier H. Pentylenetetrazole induced changes in zebrafish behavior, neural activity and c-fos expression. Neuroscience. 2005;131:759–68.
20. Afrikanova T, Serruys AS, Buenafe OE, et al. Validation of the zebrafish pentylenetetrazol seizure model: locomotor versus electrographic responses to antiepileptic drugs. PLoS One. 2013;8:e54166.
21. Balciunas D, Wangensteen KJ, Wilber A, et al. Harnessing a high cargo-capacity transposon for genetic applications in vertebrates. PLoS Genet. 2006;2:e169.
22. Zijlmans M, Jiruska P, Zelmann R, Leijten FS, Jefferys JG, Gotman J. High-frequency oscillations as a new biomarker in epilepsy. Ann Neurol. 2012;71:169–78.
23. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) method. Methods. 2001;25:402–8.
24. Macosko EZ, Basu A, Satija R, et al. Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. Cell. 2015;161(5):1202–14.
25. Ferreira TA, Blackman AV, Oyrer J, et al. Neuronal morphometry directly from bitmap images. Nat Methods. 2014;11:982–4.
26. Ristanović D, Milosević NT, Stulić V. Application of modified Sholl analysis to neuronal dendritic arborization of the cat spinal cord. J Neurosci Methods. 2006;158:212–8.
27. Baraban SC, Dinday MT, Hortopan GA. Drug screening in Scn1a zebrafish mutant identifies clemizole as a potential Dravet syndrome treatment. Nat Commun. 2015;4.
28. Dinday MT, Baraban SC. Large-scale phenotype-based antiepileptic drug screening in a zebrafish model of Dravet syndrome. eNeuro. 2015;2:8–9.
29. McNally JM, McCarley RW. Gamma band oscillations: a key to understanding schizophrenia symptoms and neural circuit abnormalities. Curr Opin Psychiatry. 2016;29:202–10.

30. Mistry AM, Thompson CH, Miller AR, Vanoye CG, George AL Jr, Kearney JA. Strain- and age-dependent hippocampal neuron sodium currents correlate with epilepsy severity in Dravet syndrome mice. Neurobiol Dis. 2014;65:1–11.

31. Hedrich UB, Liautard C, Kirschenbaum D, et al. Impaired action potential initiation in GABAergic interneurons causes hyperexcitable networks in an epileptic mouse model carrying a human Na(V)1.1 mutation. J Neurosci. 2014;34:14874–89.

32. De Lanerolle NC, Kim JH, Robbins RJ, Spencer DD. Hippocampal interneuron loss and plasticity in human temporal lobe epilepsy. Brain Res. 1989;49:387–95.

33. Robbins RJ, Brines ML, Kim JH, et al. A selective loss of somatostatin in the hippocampus of patients with temporal lobe epilepsy. Ann Neurol. 1991;29:325–32.

34. Sundstrom LE, Brana C, Gatherer M, Mepham J, Rougier A. Somatostatin- and neuropeptide Y-synthesizing neurones in the fascia dentata of humans with temporal lobe epilepsy. Brain J Neurol. 2001;124:688–97.

35. Mathern GW, Babb TL, Pretorius JK, Leite JP. Reactive synaptogenesis and neuron densities for neuropeptide Y, somatostatin, and glutamate decarboxylase immunoreactivity in the epileptogenic human fascia dentata. J Neurosci. 1995;15:3990–4004.

36. Swartz BE, Houser CR, Tomiyasu U, et al. Hippocampal cell loss in posttraumatic human epilepsy. Epilepsia. 2006;47:1373–82.

37. Houser CR. Do structural changes in GABA neurons give rise to the epileptic state? Adv Exp Med Biol. 2014;813:151–60.

38. Cobos I, Calcagnotto ME, Vilaythong AJ, et al. Mice lacking Dlx1 show subtype-specific loss of interneurons, reduced inhibition and epilepsy. Nat Neurosci. 2005;8:1059–68.

39. Yu FH, Mantegazza M, Westenbroek RE, et al. Reduced sodium current in GABAergic interneurons in a mouse model of severe myoclonic epilepsy in infancy. Nat Neurosci. 2006;9:1142–9.

40. Schoonheim PJ, Arrenberg AB, Del Bene F, Baier H. Optogenetic localization and genetic perturbation of saccade-generating neurons in zebrafish. J Neurosci. 2010;30:7111–20.

41. Sielenieks LB, Carroll NK, Van Niekerk A, et al. Evaluation of selective 5-HT2C agonists in acute seizure models. ACS Chem Neurosci. 2019;10(7):3284–95.

42. Paizanis E, Kelaï S, Renoir T, Hamon M, Lanfumey L. Life-long hippocampal neurogenesis: environmental, pharmacological and neurochemical modulations. Neurochem Res. 2007;32:1762–71.

43. Daubert EA, Heffron DS, Mandell JW, Condron BG. Serotonergic dystrophy induced by excess serotonin. Mol Cell Neurosci. 2010;44:297–306.

44. Reid CA, Leaw B, Richards KL, et al. Reduced dendritic arborization and hyperexcitability of pyramidal neurons in a Scn1b-based model of Dravet syndrome. Brain. 2014;137:1701–15.

45. Sofroniew MV, Vinters HV. Astrocytes: biology and pathology. Acta Neuropathol. 2010;119:7–35.

46. Hawkins NA, Calhoun JD, Huffman AM, Kearney JA. Gene expression profiling in a mouse model of Dravet syndrome. Exp Neurol. 2019;311:247–56.

47. Favero M, Sotuyo NP, Lopez E, Kearney JA, Goldberg EM. A transient developmental window of fast-spiking interneuron dysfunction in a mouse model of Dravet syndrome. J Neurosci. 2018;38:7912–27.

48. Hawkins NA, Zachwieja NJ, Miller AR, Anderson LL, Kearney JA. Fine mapping of a Dravet syndrome modifier locus on mouse chromosome 5 and candidate gene analysis by RNA-Seq. PLoS Genet. 2016;12(10):e1006398.

49. Kang SK, Hawkins NA, Kearney JA. C57BL/6J and C57BL/6N substrains differentially influence phenotype severity in the Scn1a+/− mouse model of Dravet syndrome. Epilepsia Open. 2019;4:164–9.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Tiraboschi E, Martina S, van der Ent W, et al. New insights into the early mechanisms of epileptogenesis in a zebrafish model of Dravet syndrome. *Epilepsia*. 2020;61:549–560. https://doi.org/10.1111/epi.16456

Chapter 6

Mechanistic multiscale modelling of cell fate

This chapter is based on my first author paper "*The Constructive Role of Noise in Cell Fate Induction*" under consideration where I developed the models and performed the analytical and simulation-based analysis. For readability and comprehensiveness, this chapter will not include the manuscript but gives a comprehensive introduction to the theoretical development and the detailed analysis.

6.1   Motivation for bottom-up approaches

The top-down analyses of the previous two chapters have revealed the immanent presence of cellular heterogeneity and indicated the importance also for clinical relevant systems. While these approaches mainly relied on the identification of pathways and molecular interactions with the aim to consistently increase knowledge in a specific context, the intrinsic mechanisms underlying cell fate dynamics as a central biological processes remained elusive. In this chapter, I address this challenge by a complementary bottom-up approach to investigate cellular heterogeneity emerging from the stochastic nature of cellular differentiation. The aim of this study is thereby not to recapitulate the details of the complex interplay of many interacting entities but to focus on a minimalist description of the underlying process to gain a more universal understanding of mechanisms at play in life as done by others [141, 142]. In particular, I investigate how the indispensable molecular noise of the gene transcription process can affect cell fate dynamics and induce cellular heterogeneity from first principles.

6.2 Overview of stochastic toggle-switch stochastic models

The starting point of my approach is the genetic toggle-switch scheme often used to describe the dynamics involved in cellular differentiation [143, 144]. The genetic toggle-switch is a simple and intuitive scheme able to reflect the most important features associated to the transcriptional regulatory mechanism of two competitive proteins. As a general idea, gene-regulatory circuits can be reconstructed from networks of simple regulatory elements and virtually reflect any desired property to reproduce a specific observed behavior [32, 145]. The toggle-switch model can be formulated by a set of Ordinary Differential Equations (ODEs) to describe the concentration evolution of gene transcription products by a mesoscopic description [146]. The corresponding stability analysis of the dynamical system allows then for the identification of steady states and the related bifurcation diagram quantifies how the nature and the number of these fixed points change in dependence on parameter values. This framework enables to uncover *critical transitions* in the dynamics, which are characterized by sudden changes of the steady state types pushing the system towards another regime [147]. Such a quick state modification is often irreversible and the system will typically never come back spontaneously to its initial state without benefiting from external energy. These abrupt transitions can mimic a sudden change of cell states or a rapid shift from health to disease conditions without the possibility to recover spontaneously. Therefore, the crucial property of a dynamical system to exhibit critical transitions might be reflect the jumps between locally stable basins of attraction representing multiple cell fates and scientific community generally call this dynamics property multistability [32, 33] .

In this context, the investigation of bistability respects the changes in regulatory mechanisms which may result in genetic switching of a bistable system. Although deterministic models have been studied to analyze the bistability properties of one system under certain conditions, they are not able to capture the cell-cell fluctuations during genetic switching. However, noise can impact the system's behavior and adding stochasticity to a deterministic model allows for investigating the role of intrinsic noise during switching of a bistable system [148]. Synthetic genetic toggle-switches have been presented as a valid tool to study the fundamental description of cellular function at DNA level in [149]. Stochastic mechanisms have been also considered of primary importance in the intention to reproduce fluctuations in regulatory signals involved into the function of transcriptional regulatory networks [55]. Specifically, the stochastic mechanisms involve molecular signals that determine when and how a given gene is transcribed. Promoter activity might also regulate the expression of other genes as schematically describe by the genetic toggle-switch. These regulatory mechanisms are intrinsically noisy due to the stochastic nature of molecular interactions of activation and inhibition, the bursts of fluctuating numbers of proteins occurring at random

time intervals and other microscopic mechanisms. The combination of these random effects leads, for instance, into phenotypic variations in isogenic population resulting in stochastic gene expression variability [150]. The work of Arkin et al. [151] also suggested that one source of the randomness expressed in the phenotype variations can be the random thermal fluctuations in the reaction rates of the chemical reactions comprising the regulatory circuity. Cellular processes are therefore intrinsically dominated by stochasticity and the deduction of key properties of cellular differentiation have also been elucidated by functional genetic modules combined with a quantitative stochastic model to capture the overall phenotypic variation by the study of noise in large-scale genetic regulatory network [152, 153]. This emergent variability is conjectured to manifest cellular heterogeneity. Both stochasticity inherent in the biochemical process of gene expression (intrinsic noise) and fluctuations in other cellular components (extrinsic noise) contribute substantially to the overall variation [86]. Transcription rates, regulatory dynamics, and genetic factors control the amplitude of noise [154]. Moreover, key properties of biochemical networks should be robust in order to ensure their proper functioning [155, 156].The kinetics of biochemical reactions involve small numbers of molecules and can be studied by exact stochastic simulation [157]. The starting point for simulating the evolution of molecular numbers is assuming a well-stirred mixture of biochemical reactions at constant temperature in a fixed volume Ω. The system contains N molecular species $\{S_1,,...,S_N\}$ that chemically interact through $M \geq 1$ reaction channels $\{R_1,,...,R_M\}$. The dynamic state of the system is denoted as $X(t) \equiv (X_1(t),...X_N(t))^T = x$ where $X_i(t)$ is the number of S_i molecules in the system at time t. For each reaction R_j, a propensity function $a_j(x)$ is defined in a given state and $a_j(x)dt$ represents the probability that one reaction R_j will occur inside Ω in the infinitesimal time interval $[t, t+dt)$. When that reaction occurs, $X(t)$ changes its state. The amount by which X_i changes is given by v_{ji}, the i-th element of the state change vector v_j which represents the change in the number of species S_i due to reaction R_j. In the discrete and stochastic case the $X(t)$ is a discrete (jump) Markov process. As such it has a time evolution equation which describes the probability $P(x,t|x_0,t_0)$ that $X(t) = x$ given $X(t_0) = x_0$ [158]. This equation is called chemical master equation (CME) and it can be written as:

$$\frac{\partial}{\partial t}P(x,t|x_0,t_0) = \sum_{j=1}^{M}(a_j(x-v_j)P(x-v_j,t|x_0,t_0) - a_j(x)P(x,t|x_0,t_0)). \tag{6.1}$$

In general, this discrete parabolic partial differential equation is too difficult to solve (either analytically or numerically) and other techniques are needed to simulate the dynamic change state. This leads to the so-called SSA of Gillespie (1977) [34], which is an statistically exact procedure for generating the time and the index of the next occurring reaction in accordance with the current values of the propensity function. There are several forms of this algorithm.

These approaches are computationally inefficient for large numbers of molecules or for high values in the propensity functions. Moreover, the reliability of these stochastic methods are based on detailed knowledge of the involved biochemical reactions, the molecular numbers and kinetic rates. However, data availability and regulatory information usually cannot provide a comprehensive picture of biological regulations [148].

In the context of cell differentiation, Strasser et al. [33] investigated gene expression stochasticity by a master equation associated to the mass action kinetics from a two-stage toggle switch model. In this approach the noise is intended to describe the intrinsic stochasticity associated to a microscopic description of gene regulatory network. Given the above described challenges of the microscopic description of biochemical networks, the detailed study of noise in large-scale genetic regulatory networks is not a realistic common approach for studying kinetic dynamics on the macroscopic description in a stochastic setting although methods have been proposed, such as stochastic Boolean models [159, 148] and probabilistic hybrid approaches [160].

To respect these challenges, my mechanistic approach to study cellular differentiation from a macroscopic description is based on the genetic toggle-switch scheme with deterministic non-linear interactions to encompass the multi-step and multi-agent interference between genotype and phenotype during molecular concentration evolution. In my proposed models, the activation and inhibition processes are described by Hill-function with coefficient equal to 1 for simplicity with the aim to reflect the biological mechanisms of self-activation and cross-inhibition macroscopically. The immanent molecular fluctuations within the cell associated to the transcription process require to consider the noise in the dynamical system. Its contribution can modulate the binary choice of cell fate in accordance with the initial conditions and the local dynamics. To investigate this phenomenon, I analyzed the model by stability analysis and investigate the effect of Gaussian noise in the context of the cell fate mechanism within a computational probabilistic framework. By using this approach, I aimed to capture the intrinsic mechanism which regulates the interface between two competitive molecules (corresponding to two alternative cell fates and/or subnetworks of genes involved in the activation of the final phenotype) and monitor the temporal evolution of protein concentrations during the continuous "struggle" occurring before a final phenotype state is established. This minimalist model approach allows for a focus on the generic mechanism also within an analytical framework and clarification of parameter dependencies. Furthermore I believe that the study of cellular differentiation in the context of a probabilistic multidimensional evolution is the appropriate choice to describe the evolution of the internal deterministic dynamics subjected to the intrinsic stochasticity and allow to eventually understand how cellular heterogeneity establishes phenotype variability.

6.3   Outline of the chapter

Throughout this chapter, I present the investigation of three versions of the model named (i) symmetric one-stage toggle-switch (T.S.) model, (ii) two-stage symmetric T.S. model and (iii) asymmetric one-stage T.S. model. In Section 6.4, I introduce a simple toggle-switch model exhibiting just one steady state as a reference. In Section 6.5 and Section Section 6.6, the first version of the model for cell fate is presented and analyzed by means of associated steady states in the general symmetric case and in its fully symmetric simplified version. In Section 6.7, I introduce the stochastic version of the cell fate model. In Section 6.8, I focused on the stationary probability distribution from a computational approach. This investigation allowed to acquire more insights into the interplay between the noise strength and initial condition and how the local proximity to other possible steady states could promote the system evolution over the stability trajectory. In Section 6.10, I computationally investigated the phenotype distributions in comparison between the one-stage T.S. model and its extension to a system with two regulatory levels of transcription and translation. To investigate the emergent behavior on the population level and how active intercellular signaling contributes to the epigenetic landscape, I added a population coupling effect to the stochastic dynamics in Section 6.10.1, which results in a positive response of the particles to the macroscopic average concentration. Finally, I present the asymmetric one-stage T.S. model as an explanatory description of the cell-fate decision triggered by an external signal that pushes the system towards a preferential attractor in Section 6.9.

6.4   A simplistic model as introduction

In this section, I introduce a simple dynamical system to describe the cross-interaction between two competing entities such as molecules or sets of genes. In this model, X and Y denote the molecular concentrations that describe a continuous system state. The temporal macroscopic evolution of X and Y is governed by the nonlinear ODEs

$$\begin{cases} \dot{X} = \frac{A}{K_X + Y} - \gamma X \\ \dot{Y} = \frac{A}{K_Y + X} - \gamma Y \end{cases} \tag{6.2}$$
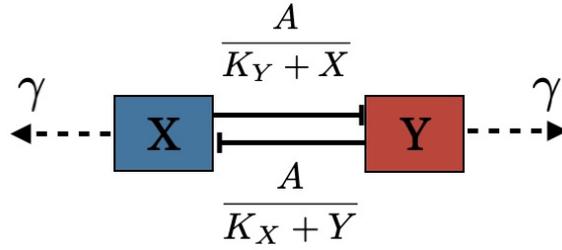
Figure 6.1: *A simple one-stage toggle switch scheme describing mutual inhibition between two competitors X and Y with degradation. The cross-inhibition terms are similar to the Hill function with coefficient equal to one.*

Fig.6.1 shows the scheme of interactions with the associated parameters where the basal rate A is set equal to the degradation rate γ for both X and Y. The non linear cross-interaction term is typically used to describe the effect of inhibition. Intuitively, in the case $K_X > K_Y$ ($K_Y > K_X$) the resulting steady state would be characterized by $Y_0 > X_0$ ($X_0 > Y_0$).

Stability Analysis By putting Eq. 6.2 equal to zero, the obtained solutions correspond to the fix points of the system. At these equilibrium concentrations, the state is classified as locally stable, unstable or saddle-node in dependence on the tendency of the system to return to, depart from or manifest a mixed direction-dependence behavior at the corresponding point in response to a local perturbation. The identification of these points was performed by subtracting the following expressions from each other:

$$\begin{cases} A - \gamma X_0 \left(K_X + Y_0 \right) = 0 \\ A - \gamma Y_0 \left(K_Y + X_0 \right) = 0 \, , \end{cases} \tag{6.3}$$

which led to the relation $Y_0 = \frac{K_X}{K_Y} X_0$. By inserting this expression into the second equation of Eqs. 6.3, the resulting equation for Y_0 exhibits only one physical solution. Hence, this minimal model obeys only the fixed point

$$\begin{cases} X_0 = \frac{K_Y}{2} \left(\sqrt{1 + \frac{4A}{\gamma K_X K_Y}} - 1 \right) \\ Y_0 = \frac{K_X}{2} \left(\sqrt{1 + \frac{4A}{\gamma K_X K_Y}} - 1 \right) \end{cases} \tag{6.4}$$

.

The nature of the state follows from the Jacobian Matrix at this point by the sign of the two corresponding eigenvalues. If the real part of the biggest eigenvalue is negative, the state is defined as stable; if the real part of the biggest eigenvalue is positive, the state is unstable, and in case one of them is zero then the steady state is a saddle-node [161]. By using *Mathematica*, I verified that the steady state (X_0, Y_0) is a stable state for all positive parameter values. Since this simple model has only one attractor state it is not an appropriate model to describe the binary cell fate decision mechanism. A stem cell can potentially end up at several different stable states which are equally likely at the beginning of the differentiation process. During this process, the cells acquire more and more specificity by approaching in parallel one of the possible cell states. Thereby at least bistability is necessary to investigate the mechanism of switching between two phenotypic traits. A toggle-switch genetic motif exhibiting multistability in the dynamics is therefore a minimum request to be an explanatory model of the process [162]. For this purpose, I expanded the above minimalist model by introducing a multiplicative term accounting for the interplay between self-activation and cross-inhibition regulations in the following section.

6.5   Symmetric one-stage toggle switch model

The toggle-switch motif with its underlying dynamics for self-activation and cross-inhibition became a popular approach to model binary cell-fate decisions as a balance of two lineage-affiliated transcription factors [116]. In this respect, the model developed and analysed here is based on the gene-circuit scheme in the attempt to capture basic mechanisms emerging from the interaction between non-linear dynamics and stochasticity without intending to study specific activation and repression pathways. Here, the genetic toggle switch is not intended in its classical application to model binary cell-decision. Typically, the toggle switch is used in a configuration, which exhibits bistability or multistability to study the selection of steady states and how stochasticity interferes with this process at the same scale of the involved biomolecular reaction [148] by analysing the noise-triggered movement of the system on the phase space [33]. In the perspective of this study, the toggle switch scheme is introduced to model a more general mechanism of cell induction as a spontaneous appearance from the combined dynamical interfere between noise and deterministic instruction.

The parameters A and γ in Fig. 6.2 indicate the rate of molecular production of e.g. protein or mRNAs, and the parameters K_X and K_Y modulate the swapping between X and Y according the underlying non-linear dynamics. Thereby, the proposed dynamics is not intended to represent a preexisting gene program that is adapted and subsequently adopted by the cells at the commitment event, but it is meant to mechanistically describe (dynamically) distinct states preceding the final cell fate choice. Specifically, the

symmetric one-stage toggle switch reflects the putative cell behavior in the transitory metastable state and its asymmetric version addresses the fast and non-linear dynamical changes of the internal gene rearrangement upon stimuli eventually conducting to the final stable state 2.1. In this perspective, the external triggering signal towards a specific lineage specification is not modeled as a critical parameter, which would induce a critical transition in the dynamics, but it modifies significantly the internal structural gene dynamics. In the here presented model, the asymmetry between K_X and K_Y describes the stage where cells are forced to an immanent choice between two unexpected stable cell states due to a rapid and definite change of the phase space structure. The main finding of this analysis is that the noise contribution allows for a more reliable identification of the induced cell state and extends the deterministic dynamical evolution within the biology distributions context 3.8.

In the model, a self-interaction term was added in form of a Hill activation function (with Hill coefficient equal to 1) similar to other approaches in literature [163, 164]. The scheme of the model is shown in Fig. 6.2.
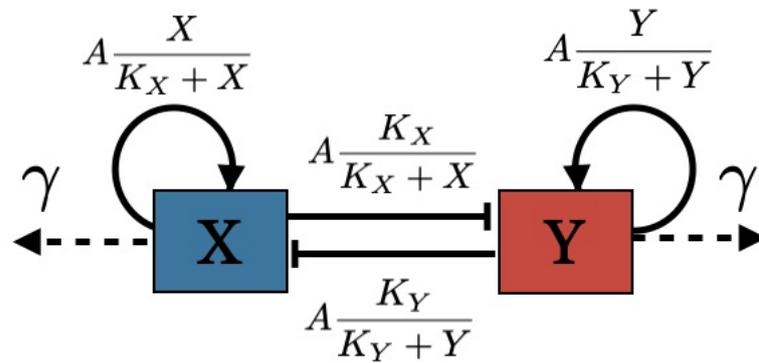
Figure 6.2: *One-stage toggle switch scheme describing mutual inhibition between X and Y accompanied by degradation and self-activation depending on the concentration according to the Hill function for the activators. Analogously, the cross-inhibition terms are similar to the Hill function for repressors. For both contributions the Hill coefficient is equal to one.*

The resulting non-linear coupled differential equations read

$$\begin{cases} \dot{X} = A \frac{K_X}{K_X+Y} \frac{X}{K_X+X} - \gamma X \\ \dot{Y} = A \frac{K_Y}{K_Y+X} \frac{Y}{K_Y+Y} - \gamma Y \, , \end{cases} \tag{6.5}$$

where A is the maximum production rate, K_X and K_Y correspond to the activation and mutual repression coefficients of X and Y, respectively. Hence, the concentrations of the two competitive molecules decrease with the increase of the other one by respecting the Hill function for a repressor. I decided to insert a non-linear term as a product between activation and inhibition contribution to macroscopically enclose the

intrinsic non-linear molecular response to opposite activation signals of transcription factors. Moreover, for simplicity I chose the same degradation rate γ for both variables.

6.5.1 Steady state classification

The steady states were identified by evaluating the dynamics given by Eq. 6.5 when the concentrations are locally constant under a small perturbation in analogy to Section 6.4. The stationary concentrations $X^{(0)}$ and $Y^{(0)}$ follow the solution of the system

$$\begin{cases} AK_X X^{(0)} - \gamma X^{(0)} \left(K_X + X^{(0)}\right)\left(K_X + Y^{(0)}\right) = 0 \\ AK_Y Y^{(0)} - \gamma Y^{(0)} \left(K_Y + Y^{(0)}\right)\left(K_Y + X^{(0)}\right) = 0 \,. \end{cases} \tag{6.6}$$

The trivial solution is the steady state $\bar{X}_1 = (0,0)$. The first not-trivial steady state was found by imposing $X^{(0)} = 0$ and $Y^{(0)} \neq 0$ leading to $\bar{X}_4 = \left(0, \frac{A}{\gamma} - K_Y\right)$. Another possible steady state is equivalently $\bar{X}_5 = \left(\frac{A}{\gamma} - K_X, 0\right)$. To ensure that both points are physical solution, the inequality $\frac{A}{\gamma} \geq max\{K_X, K_Y\}$ has to be hold. The properties that $X^{(0)} \neq 0$ and $Y^{(0)} \neq 0$ allowed to divide both equations respectively by $X^{(0)}$ and $Y^{(0)}$ and by subtracting from the first equation the second one, the following relation occurs

$$A\left(K_X - K_Y\right) - \gamma\left(\left(K_X + K_Y\right)\left(K_X - K_Y\right) + \left(X^{(0)} + Y^{(0)}\right)\left(K_X - K_Y\right)\right) = 0 \,. \tag{6.7}$$

By defining $\epsilon \equiv K_X - K_Y$ and dividing Eq. 6.7 by ϵ, the dependency on the difference $K_X - K_Y$ disappears and lead to a linear relation between $X^{(0)}$ and $Y^{(0)}$:

$$X^{(0)} = \frac{A}{\gamma} - K_X - K_Y - Y^{(0)} \,, \tag{6.8}$$

which has to be positive for a physical solution and eventually leads to the maximum concentration for $Y^{(0)}$ as

$$Y^{(0)} \leq Y_{max} \equiv \frac{A}{\gamma} - K_X - K_Y. \tag{6.9}$$

By substituting Eq. 6.8 into the first equation in Eqs. 6.6, we obtain a second order equation for $Y^{(0)}$ as

$$Y^{(0)^2} + \left(K_X + K_Y - \frac{A}{\gamma}\right)Y^{(0)} + K_X K_Y = 0. \tag{6.10}$$

The solutions for this stationary Y concentrations referring to the two steady states $\bar{X}_2$ and $\bar{X}_3$ obey the

relation

$$Y_{\pm}^{(0)} = \frac{Y_{max}}{2} \pm \frac{\sqrt{Y_{max}^2 - 4K_X K_Y}}{2} \tag{6.11}$$

$$= \frac{1}{2}\left(\frac{A}{\gamma} - K_X - K_Y\right) \pm \frac{1}{2}\sqrt{\left(K_Y + K_X - \frac{A}{\gamma}\right)^2 - 4K_X K_Y},$$

and the corresponding stationary X concentrations are given by

$$X_{\mp}^{(0)} = \frac{Y_{max}}{2} \mp \frac{\sqrt{Y_{max}^2 - 4K_X K_Y}}{2} \tag{6.12a}$$

$$= \frac{1}{2}\left(\frac{A}{\gamma} - K_X - K_Y\right) \mp \frac{1}{2}\sqrt{\left(K_Y + K_X - \frac{A}{\gamma}\right)^2 - 4K_X K_Y} \tag{6.12b}$$

$$\equiv Y_{\mp}^{(0)}. \tag{6.12c}$$

The steady state configuration is therefore given by a specific concentrations pair (X, Y) with $\bar{X}_2 \equiv \left(Y_+^{(0)}, Y_-^{(0)}\right)$ and $\bar{X}_3 \equiv \left(Y_-^{(0)}, Y_+^{(0)}\right)$ where $Y_+^{(0)} + Y_-^{(0)} = Y_{max}$ holds. Note that the steady solutions are invariant under the exchange between K_X and K_Y, meaning that the steady states remain the same. It confirms the equivalent role of X and Y as expected for the symmetric configuration dynamics. I further simplified the model by setting $K_X = K_Y$ without loosing generality since both variables level out the antagonistic effect of the other variable by the self-activation action.

6.6 Deterministic analysis of the fully symmetric model

To further investigate the model properties, I next investigated analytically the dynamics of the fully symmetric version of the one-stage symmetric T.S. model by putting $K_X = K_Y = K$. By choosing the equivalence between K_X and K_Y parameter, I focused on the intrinsic mechanism that regulates the interaction between X and Y with the same strength in regulating the self-activation and cross-inhibition activity. The corresponding scheme is shown in Fig. 6.3.
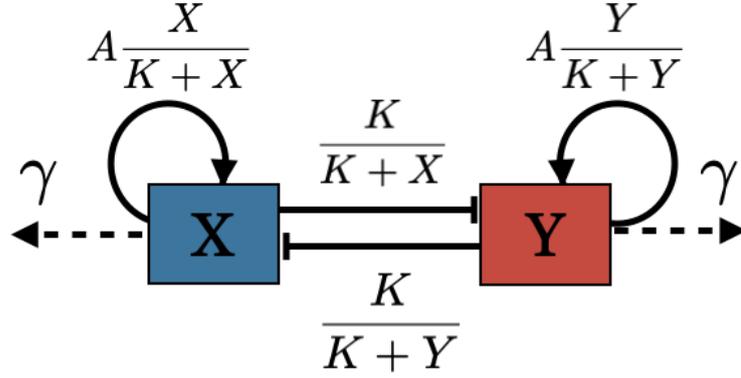
Figure 6.3: *One-stage toggle switch scheme describing mutual inhibition between X and Y accompanied by protein degradation and self-activation. This scheme is a fully symmetric description of the symmetric one-stage T.S. model since X and Y operate with the same parameter setting.*

The steady states for this fully symmetric model are given by

$$\bar{X}_1 = (0,0) \tag{6.13a}$$

$$\bar{X}_2 = \left(\frac{1}{2}\left(\frac{A}{\gamma} - 2K\right) + \frac{1}{2}\frac{A}{\gamma}\sqrt{1 - \frac{4K\gamma}{A}}, \frac{1}{2}\left(\frac{A}{\gamma} - 2K\right) - \frac{1}{2}\frac{A}{\gamma}\sqrt{1 - \frac{4K\gamma}{A}}\right) \tag{6.13b}$$

$$\bar{X}_3 = \left(\frac{1}{2}\left(\frac{A}{\gamma} - 2K\right) - \frac{1}{2}\frac{A}{\gamma}\sqrt{1 - \frac{4K\gamma}{A}}, \frac{1}{2}\left(\frac{A}{\gamma} - 2K\right) + \frac{1}{2}\frac{A}{\gamma}\sqrt{1 - \frac{4K\gamma}{A}}\right) \tag{6.13c}$$

$$\bar{X}_4 = \left(0, \frac{A}{\gamma} - K\right) \tag{6.13d}$$

$$\bar{X}_5 = \left(\frac{A}{\gamma} - K, 0\right). \tag{6.13e}$$

The physical solutions with the constrain for the concentrations being always a positive real number, imposes a maximum value for K. Specifically, the existence of a real square root in $\bar{X}_2$ and $\bar{X}_3$ requires the condition $K \leq \frac{A}{4\gamma} \equiv K_{max}$ and positive values for the non-null component of $\bar{X}_4$ and $\bar{X}_5$ concentrations imposes the condition $K \leq \frac{A}{\gamma} = 4K_{max} \equiv K_{max}^{(2)}$. Thus, the steady states can be expressed in term of K_{max}

as

$$\bar{X}_1 = (0,0) \tag{6.14a}$$

$$\bar{X}_2 = \left(2K_{max}\left[1 - \frac{K}{2K_{max}} + \sqrt{1 - \frac{K}{K_{max}}}\right], 2K_{max}\left[1 - \frac{K}{2K_{max}} - \sqrt{1 - \frac{K}{K_{max}}}\right]\right) \tag{6.14b}$$

$$\bar{X}_3 = \left(2K_{max}\left[1 - \frac{K}{2K_{max}} - \sqrt{1 - \frac{K}{K_{max}}}\right], 2K_{max}\left[1 - \frac{K}{2K_{max}} + \sqrt{1 - \frac{K}{K_{max}}}\right]\right) \tag{6.14c}$$

$$\bar{X}_4 = \left(0, 4K_{max} - K\right) \tag{6.14d}$$

$$\bar{X}_5 = \left(4K_{max} - K, 0\right). \tag{6.14e}$$

To characterize in more detail the non-linear dynamics and to verify the presence of the obtained steady state solutions, I investigated the vector field of the phase space where each point corresponds to the fixed initial conditions leading to a 2-dimensional vector of the finite increments (Δ_X, Δ_Y) at any point on the underlying grid. The resulting arrows indicate the instantaneous direction of the concentration changes in the phase space as shown for $K = 20$, $K = 25$, $K = 30$, $K = 50$ and $K = 80$ in Fig. 6.4. The result for $K = 20$ (Fig. 6.4a) confirms the presence of all the identified steady states where $\bar{X}_1$ is the trivial unstable state, $\bar{X}_2$ corresponds to the pink point and $\bar{X}_3$ to the red point, respectively, and the blue and green point indicate the steady states $\bar{X}_4$ and $\bar{X}_5$. Interestingly, the arrow flow converge to a line of steady points. On this trajectory in the phase space, the system is in a stationary condition. The points $\bar{X}_2$ and $\bar{X}_3$ are on the extreme value of the two parts of the trajectory that is highlighted by the orange and pink lines. These two parts of the steady points can be analytically found by analyzing the relation between $\bar{X}_2$ and $\bar{X}_3$. Indeed, they emerge from the $\bar{X}_2$ and $\bar{X}_3$ formulas (Eqs. 6.14 and 6.12a), evaluated for $20 < K < K_{max}$, in the following way:

$$Y_+^{(0)} - Y_-^{(0)} = 2K_{max}\sqrt{1 - \frac{K}{K_{max}}}, \tag{6.15}$$

which is the analytical hyperbolic curve corresponding to the drawn dashed line on the vector fields. The degeneration of these steady states is explainable by the symmetry of the dynamics as the X (Y) molecular concentration component for the $\bar{X}_1$ corresponds to Y (X) component for $\bar{X}_2$.

The steady states $\bar{X}_4$ and $\bar{X}_5$ are not on the stability line but are very close to it. This might be a result of the specific symmetric case in the regime $K < K_{max}$ which encloses the presence of another stability configuration of the system. In the Fig. 6.4b, the vector field for $K = K_{max} = 25$ is shown. In this case, the stationary concentrations $\bar{X}_2$ and $\bar{X}_3$ coincide and have the component X and Y equal each other. Although this extreme case is characterized by three steady states, the trajectory (dashed line) remains as

71

the stability manifold of the system corresponding to the collection of steady states $\bar{X}_2$ and $\bar{X}_2$ evaluated for $0 < K \leq K_{max}$.

For the regime $K > K_{max}$, the stationary trajectory is expected to disappear but in Figure 6.4c the corresponding vector field for $K = 30$ still indicates a stationary trajectory of the system. The dashed line corresponds to the stability trajectory in the regime $K < K_{max}$. The analysis for $K = 50$ in Fig. 6.4d exhibits more clearly how the system changes its stability line. The dashed line is still referred to the previous stability trajectory to highlight the completely disappearance of the previous line of steady points. In this configuration, the system was found to have again a stability line going through the steady states $\bar{X}_4$ and $\bar{X}_5$. Therefore the choice of $K = 50$ puts the system in another stability regime, which is imposed by the full symmetry in the molecular concentrations of the alternative possible steady states. The vicinity of the steady states around $K = K_{max}$ masks these distinct regimes since both co-exists during the transition from one stable phase to the other one, appearing as a degeneration of the fully symmetric regime.
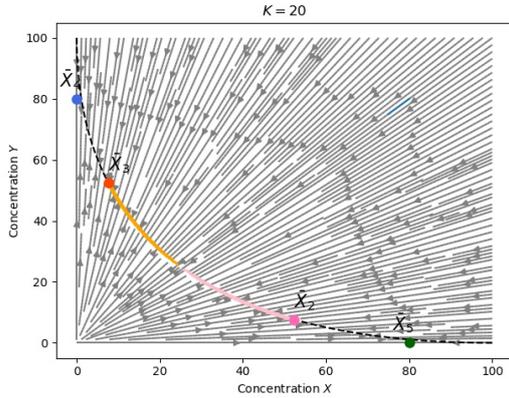
(a) *Vector field of the dynamics for $K = 20$*
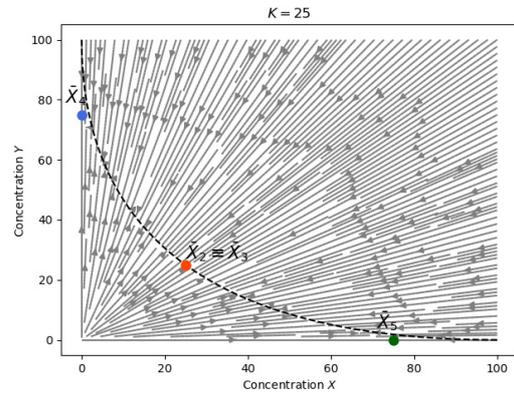
(b) *Vector field of the dynamics for $K = K_{max} = 25$*
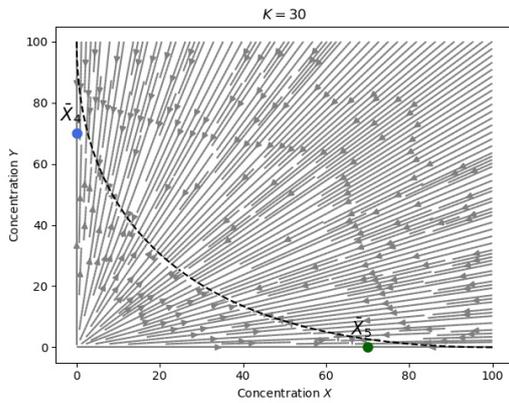
(c) *Vector field of the dynamics for $K = 30$*
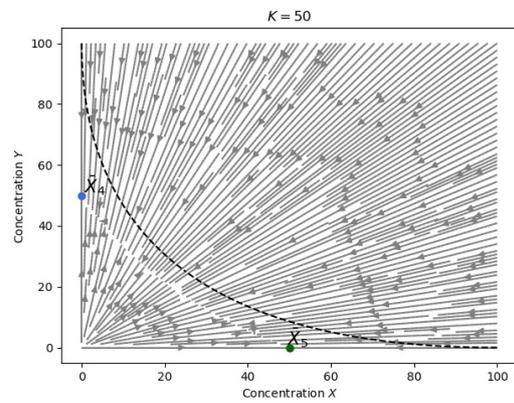
(d) *Vector field of the dynamics for $K = 50$*

Figure 6.4: *Stationary points and vector field for the fully symmetric model. The colored points denote the steady states: $\bar{X}_2$, $\bar{X}_3$, $\bar{X}_4$ and $\bar{X}_5$ by the pink, red, blue and green dots, respectively. In Fig. 6.4b the steady points $\bar{X}_2$, $\bar{X}_3$ coincide at the red point. The dashed line connects all the other possible $\bar{X}_2$ and $\bar{X}_3$ obtainable in the regime $0 < K < 25$. In Fig. 6.4c, $\bar{X}_2$, $\bar{X}_3$ are not physically acceptable since we are in the regime $K > K_{max}$ but the entire stationary trajectory is still accessible from the system. The steady points $\bar{X}_4$ and $\bar{X}_5$ are just slightly diverging from the stability trajectory defined by Eq. 6.15. Fig.6.4d highlights further the existence of a steady state trajectory, which connects $\bar{X}_4$ with $\bar{X}_5$.*

Finally, by increasing K to 80, the associated vector field is shown in Fig.6.5 and the alternative stability line is more evident and corresponds to the straight line that crosses $\bar{X}_4$ and $\bar{X}_5$. It might correspond to the degeneration of the hyperbolic curve since in this new regime the two steady states are well-defined but still equally preferred from the system.

73
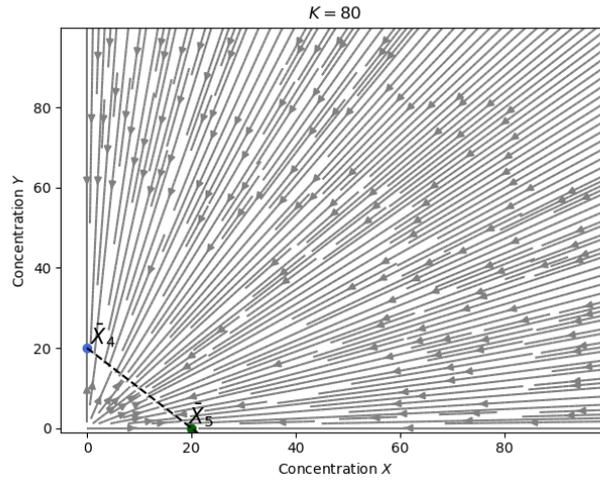
Figure 6.5: *Vector field of the dynamics of the fully symmetric case $K_X = K_Y = 80$. The stability trajectory is degenerate into a linear line which connects the two steady points $\bar{X}_4$ and $\bar{X}_5$.*

For a systematic analysis of the potential steady states, the corresponding bifurcation diagram was calculated with K as control parameter. For this purpose, Fig. 6.6 shows the stationary X and Y concentrations as function of K where the full symmetry of the dynamics induces the switch within the pair $\bar{X}_2$ and $\bar{X}_3$ and the pair $\bar{X}_4$ and $\bar{X}_5$ (see color code switching in Figs. 6.6a and 6.6b). The analysis shows also that K_{max} is the maximum value to have distinguishable steady points associated to $\bar{X}_2$ and $\bar{X}_3$.

(a) *X-concentration defining state*

(b) *Y-concentration defining state*
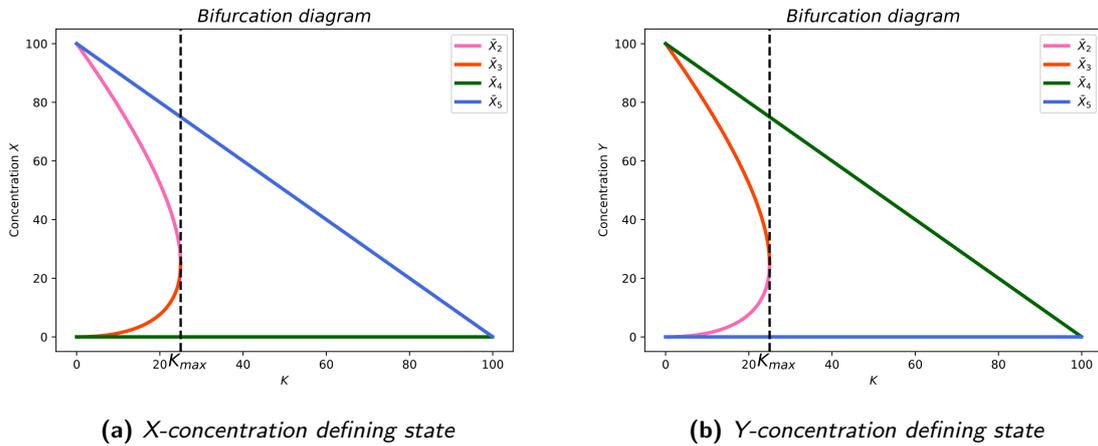
Figure 6.6: *Bifurcation diagrams with control parameter K as a collection of all accessible steady states for the system identified by X and Y concentrations in the left and right panel, respectively. Note that the full symmetry induce the swapping of steady states for X and Y indicated by the colors.*

74

This analysis indicates that the fully symmetric case $K_X = K_Y = K$ is associated to a degeneration of the stable manifold where the final steady states of the system depends on the initial conditions due to non-linearity of the dynamics. Thus, the system is able to reach potentially any possible steady state configuration established by any K parameter in a defined regime. Furthermore, in the vicinity of the fixed points in the phase space $\bar{X}_4$ and $\bar{X}_5$, the stationary trajectory does not avoid the system to reach the degenerated steady states. Therefore, the system exhibits a sort of adaptability in its dynamical behaviour to e.g. react to external signals. Since this is an essential property for the robustness of cell fate dynamics, we further investigated the nature of these steady states.

6.6.1 Steady state classification

The nature of the identified fixed points in Section 6.14 has been uncovered systematically by evaluating the sign of the two associated eigenvalues of the Jacobian matrix of the dynamical system at the corresponding steady states. The Jacobian matrix takes the general form

$$J(A, \gamma, K\}) = \begin{pmatrix} \frac{AK^2}{(K+Y)(K+X)^2} - \gamma & \frac{-AKX}{(K+X)(K+Y)^2} \\ \frac{-AKY}{(K+Y)(K+X)^2} & \frac{AK^2}{(K+X)(K+Y)^2} - \gamma \end{pmatrix}$$

At $\bar{X}_1$, the eigenvalues are identical and both positive or null in the allowed regime $K \leq \frac{A}{\gamma}$. Hence, this point is an unstable state for any K except for $K = \frac{A}{\gamma}$ with zero eigenvalues where the point changes into a limit cycle. For $\bar{X}_4$ and $\bar{X}_5$, one eigenvalue is always zero and the other is always negative in the stable regime, except at $K = \frac{A}{\gamma}$ where again the points become a limit cycle because both eigenvalues are vanishing. On the eigenspace associated to the eigenvalue equal to zero, the solutions are therefore time independent due to vanishing Lyapunov exponents indicating the independence on external perturbations and therefore the eigenspace is a line of equilibria. The presence of the attractive line of equilibria makes the point stable but not asymptotically stable for $K \neq \frac{A}{\gamma}$.

The analytical solution for the eigenvalues of the points $\bar{X}_2$ and $\bar{X}_3$ are not straight forward, and therefore a *Python* script was used to evaluate the nature of the points for K across the range between 0 and $K_{max}^{(2)}$ defining the region of stability as shown in Fig. 6.7.
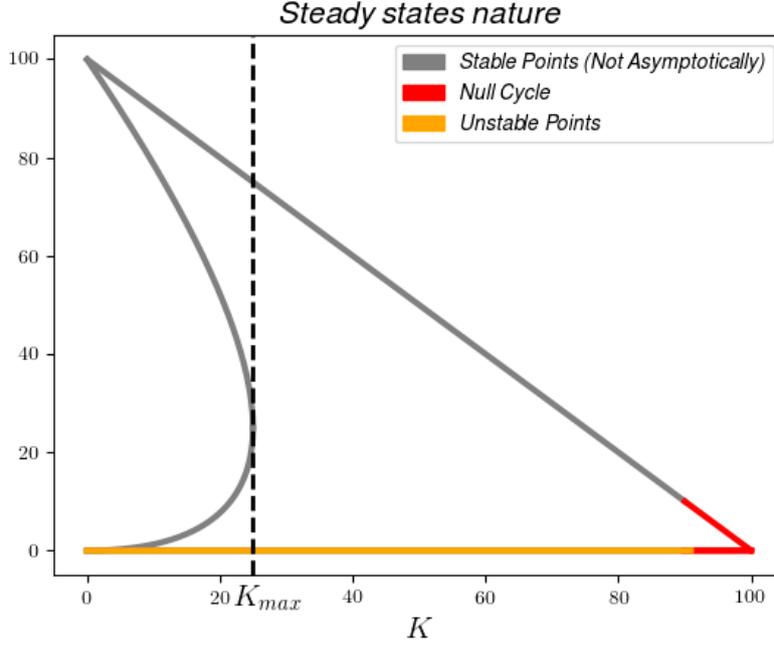
Figure 6.7: *Bifurcation diagram for the steady states related to one-stage toggle-switch model in the fully symmetric case $K_X = K_Y = K$.*

The result shown in Figure 6.7 indicates that for all K the steady states associated to $\bar{X}_2$, $\bar{X}_3$ and $\bar{X}_5$ are (not asymptotically) stable states because of the presence of one vanishing eigenvalue. The state $\bar{X}_0$ is always an unstable point. For $K \approx K_{max}^{(2)}$ the steady states become null cycles, characterized by the vanishing of both eigenvalues. The steady states $\bar{X}_2$ and $\bar{X}_3$ collapse to one when $K \equiv K_{max}$ corresponding to $\left(\frac{A}{4\gamma}, \frac{A}{4\gamma}\right)$ concentration values (6.14b,6.14c) . This point is the extreme value corresponding to the boundary of the fully symmetric case and it might be illuminating to investigate its peculiar role in the separation of two distinct regimes in detail. For this purpose, I specifically analyzed the interplay between non-linear dynamics and stochastic effects around this point to investigate the system's dynamics including the appearance of potential transitions.

6.7 Stochastic analysis of the model

In this section, I present the stochastic study of the one stage symmetric toggle-switch model by computational results.
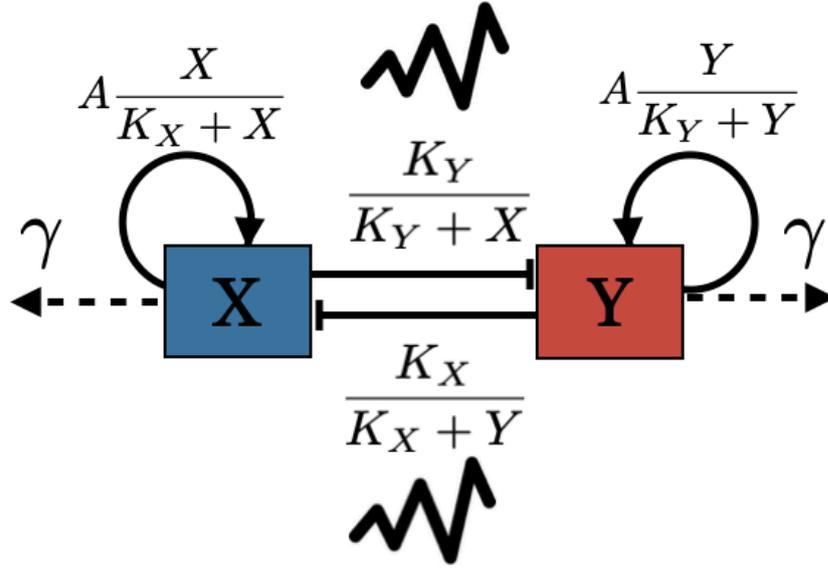
Figure 6.8: *Representation of the stochastic one-stage toggle switch model describing mutual inhibition between X and Y accompanied by degradation and self-activation depending on the corresponding concentrations by the Hill function for activators. Equivalently, the cross-inhibition terms are similar to the Hill function for repressors. For both contributions, the Hill coefficient is set to one.*

The stochastic version of the ODEs system 6.5 takes the form

$$
\begin{cases}
\dot{X} = \frac{AK_X}{K_X+Y} \frac{X}{K_X+X} - \gamma X + \eta_X(t) \\
\dot{Y} = \frac{AK_Y}{K_Y+X} \frac{Y}{K_Y+Y} - \gamma Y + \eta_Y(t)
\end{cases}
\tag{6.16}
$$

where η_X and η_Y denote the random forces representing the effect of the intracellular active processes. These forces have a Gaussian probability distribution with correlation function:

$$
\langle \eta_i(t)\eta_j(t') \rangle = 2D_i \delta ij \delta(t-t')
\tag{6.17}
$$

where D_i with $i = (X, Y)$ is the diffusion coefficient. The stochastic terms contribute to the concentration changes during a time interval bigger than the temporal variation associated to the dynamics itself. Indeed, the X and Y molecular concentration, which define the cell state in our model, are slow variables since they require more time to relax to equilibrium compared to the local stochastic fluctuations. In the following analysis, computational simulations have been performed to get insights into the interplay between the non-linearity and the diffusive effect interfere.

A mathematical approach to study the probabilistic evolution of the system is by writing the corresponding Fokker-Plank equation (FPE) [165] describing the temporal evolution of the X and Y concentrations as random variables subjected to a drift term and the Gaussian white noise. While the general analytical solution of FPEs are often not achievable, computational study provide important insights into the property of stochastic systems.

6.8   Stationary solution of FPE

The stationary solution of the corresponding FPE has been computationally analyzed. Specifically, in this section, computational investigations of the stationary probability distributions allowed to study the interplay between noise and the deterministic non-linear dynamics. The stationary probability distribution of a system is defined locally as temporally stable and is in principle strictly independent on the initial condition of the dynamics. In this section, the stationary solution of the FPE describing the fully symmetric toggle-switch model is studied in several regions of the phase space. The stability manifold, which is composed by all admissible steady states is supposed to be the domain of the corresponding stationary probability distribution ($D_X = D_Y$ is hold for simplicity in the computational simulations). Note that the points on the stability manifold can be found by evaluating the stationary molecular concentrations for any values of K. Moreover, by focusing on the regime $K \leqslant K_{max}$, the steady states solutions define two distinct branches on the bifurcation diagram as shown in Fig. 6.6.

In order to visualize the probability distribution at the stationary profile in detail, corresponding stochastic simulations were performed in the *Julia* programming environment. The system was simulated with parameters set as $K = 20$, $A = 0.5$, $\gamma = 0.005$, $D = 0.0001$ for a total simulation time period of $T = 10000$ ($T = 8000$ is roughly the required time from the system to reach a stationary configuration starting at the chosen initial condition) with a time step of $dt = 0.1$. To respect the stochastic dynamics, a number of $M = 70,000$ trajectories was modelled with the initial position in phase space defined by $X_0 = 50$ and $Y_0 = 9$. The result of the simulation is visualized as a 2-dimensional histogram on the (X, Y) phase space and shows the concentrations at the final time T in Fig.6.9a where the color map indicates the graded normalized value of the trajectories density in the (X, Y) space of molecular concentrations.

(a) (X,Y) space representation   **(b) (κ,ζ) space representation**
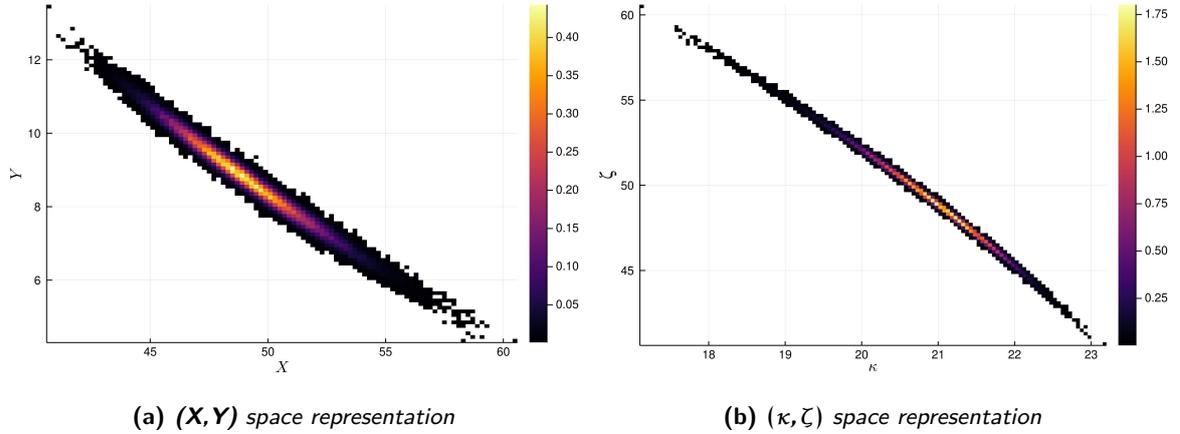
Figure 6.9: *Probability distribution for an ensemble of 70,000 cells starting from the initial condition $(X_0, Y_0) = (50, 9)$. 2-dimensional histograms of the probability distribution for an ensemble of 70,000 cells starting from the initial condition $(X_0, Y_0) = (50, 9)$ and gathered at the final time $T = 10000$, when the system is in a stationary configuration, with dynamical parameters $K = 20$, $A = 0.5$, $\gamma = 0.005$ and $D = 0.0001$.*

To computationally investigate the property of the stationary probability distribution, a non-linear variable change was performed in the representation space of the 2-dimensional histogram of the probability distribution. By exploiting the $\frac{X-Y}{2}$ quantity from Eq. 6.14b, a new variable κ is introduced for stationary concentrations of the steady state $\bar{X}_2$. The variable change in the representation space of the computed probability distribution is defined by

$$\kappa = K_{max} - \frac{(X-Y)^2}{16K_{max}} \, , \tag{6.18}$$

which can subsequently be used for a new set of variables $(\kappa, X) \equiv (\kappa, \zeta)$ to visualize the result of the stationary probability distribution of the computational simulations shown in Fig. 6.9b.

Interestingly, the probability distribution in this new representative space was found to be mainly concentrated on the stability manifold. Indeed, the branch referred to the steady states $\bar{X}_2$ emerges as the domain of the stationary probability in the (ζ, κ) space. Hence, the new set of variables is therefore an appropriate representation for the stationary molecular distributions. The shown stationary distribution is located along the stability line and peaked around $(X, Y) = (20.5, 49)$. By investigating the the vector field in Fig. 6.4a, it becomes evident that the system is following the dynamics for the specific initial condition towards the corresponding final steady state on the branch. This might indicate that the effect of the noise on the dynamics is to spread the stationary probability distribution along the stationary manifold and locates at the final point defined by the vector flux.

To support this hypothesis, another configuration was simulated for which the other branch of the

stability trajectory was expected as the domain of the resulting stationary probability distribution in the (ζ, κ) space. For the same parameter setting with $A = 0.5$, $\gamma = 0.005$ and $D = 0.0001$, the initial condition for the ensemble was set to $(X_0, Y_0) = (7, 52)$. The resulting dynamics is plotted on the molecular concentration space in Fig. 6.10a.

(a) *(X,Y)* space representation   (b) (κ, ζ) space representation
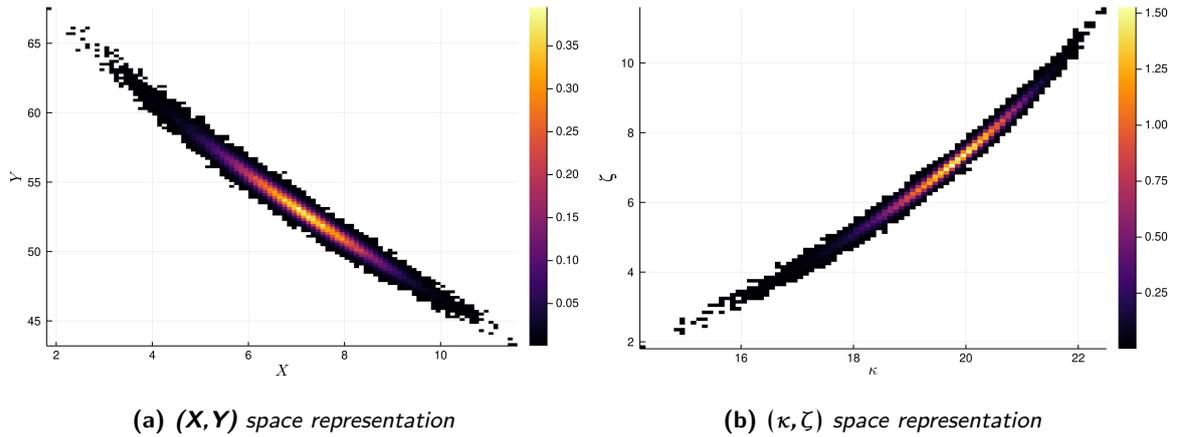
Figure 6.10: *Stationary probability distribution for an ensemble of 70,000 samples starting from the initial condition $(X_0, Y_0) = (7, 52)$. 2-dimensional histograms of the stationary probability distribution for an ensemble of 10,000 samples starting from the initial condition $(X_0, Y_0) = (7, 52)$ and gathered at the final time T=10,000. The system has the dynamical parameters: $K = 20$, $A = 0.5$, $\gamma = 0.005$ and $D = 0.0001$.*

After the change from the (X, Y) space into the new space of (κ, ζ), the 2-dimensional histogram reflects the general behavior of the system to end up at a precise steady state on the stability manifold specified by its initial condition as shown in Fig. 6.10b. Thereby the noise diffuses the probability distribution on a region around the alternative branch of the stability trajectory as expected. As another case, the extreme configuration with $K = K_{max}$ was investigated to confirm that the stationary probability distribution is likely to be centered at $(X, Y) = (K_{max}, K_{max})$ and to quantify the spreading around this point. Based on the vector field analysis in Fig. 6.4b, the initial condition was set to $X_0 = Y_0$ as the preferential direction towards the extreme point. The results of the computational simulation are shown in Figs. 6.11 and 6.12, respectively.
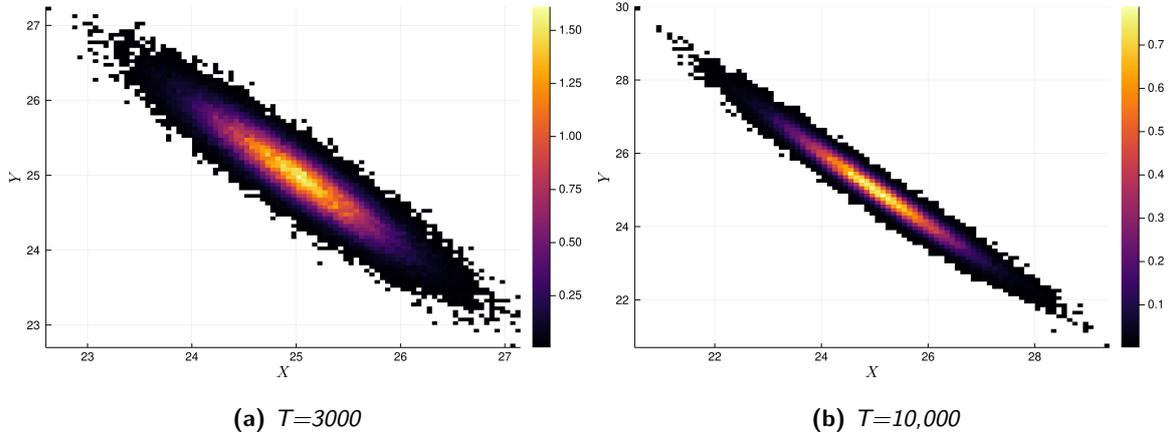
(a) *T=3000*

(b) *T=10,000*

Figure 6.11: *Probability distributions at the intermediate simulation time T=3000 and the final time T=10,000 which corresponds to the stationary probability profile. The probability distributions are represented as a 2-dimensional histogram on the (X,Y) space from a statistical ensemble of 70,000 samples. The used parameters are $A = 0.5$, $\gamma = 0.005$, $K = 25$ and $D = 0.0001$ with the initial condition $(X_0, Y_0) = (50,50)$.*

(a) *T=3000*

(b) *T=10,000*
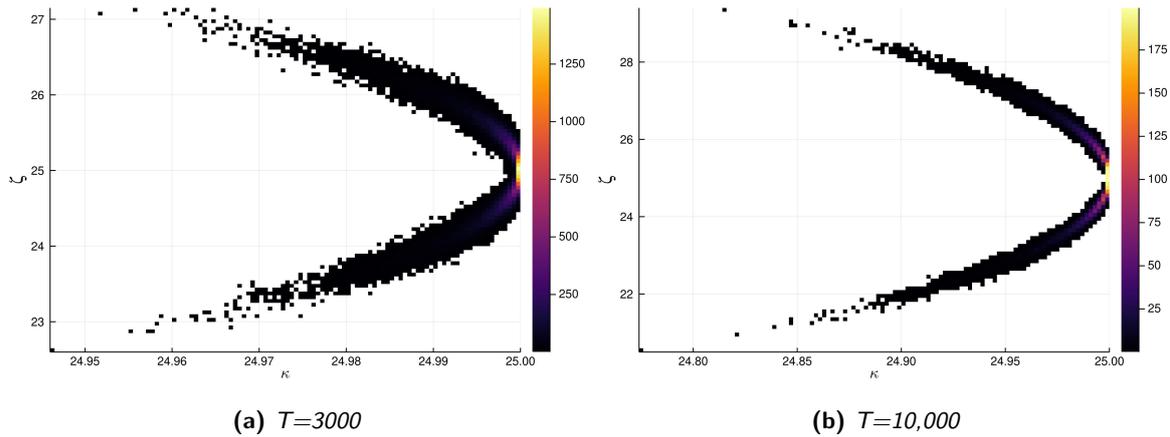
Figure 6.12: *Probability distributions at the intermediate simulation time T=3000 and the final time T=8000 which corresponds to reaching a stationary probability profile. The probability distributions are represented as a 2-dimensional histogram on the in the (κ,ζ) space from a statistical ensemble of 70,000 samples. The parameters were set to $A = 0.5$, $\gamma = 0.005$, $K = 25$ and $D = 0.0001$ with the initial condition $(X_0, Y_0) = (50,50)$.*

For the extreme case with $K = K_{max}$, the appearance of the two branches of the stability trajectory is clearly visible in the (κ,ζ) space similar to the regime for $K \leq 25$. The noise is here the key element to give to the system the ability to explore a portion of the stability trajectory according to the initial condition. The system can explore other possible steady states only due to the stochastic nature of the dynamics. Moreover, the peak of the distribution in Fig.6.12 corresponds to $\kappa = 25$ and is in agreement with the

81

vector field analysis in Fig. 6.4b. Indeed, the system will end up to reach the steady state (K_{max}, K_{max}) for any initial points on the $X_0 = Y_0$ line. This is a particular point since from there the noise spreads the stationary probability distribution on both branches. To complete the investigation, one configuration in the regime $K > 25$ was studied by setting $K = 80$ and the initial condition to $(25, 15)$ based on the vector field analysis in Fig. 6.5. The result highlights again the way of the probability distribution to asymptotically reach a precise steady state according to the initial condition as shown in Fig.6.13a. In this case, the stability line connects the steady state points defined by Eqs. 6.14d and 6.14e in agreement with the stability analysis in Section 6.6. Since for these states varying K leads to a movement along the lines $X = 0$ and $Y = 0$, respectively, the linear dependency cannot be recovered in contrast to the regime $K \leq K_{max}$.

(a) *(X,Y) space representation*   **(b)** *(X,Y') space representation*
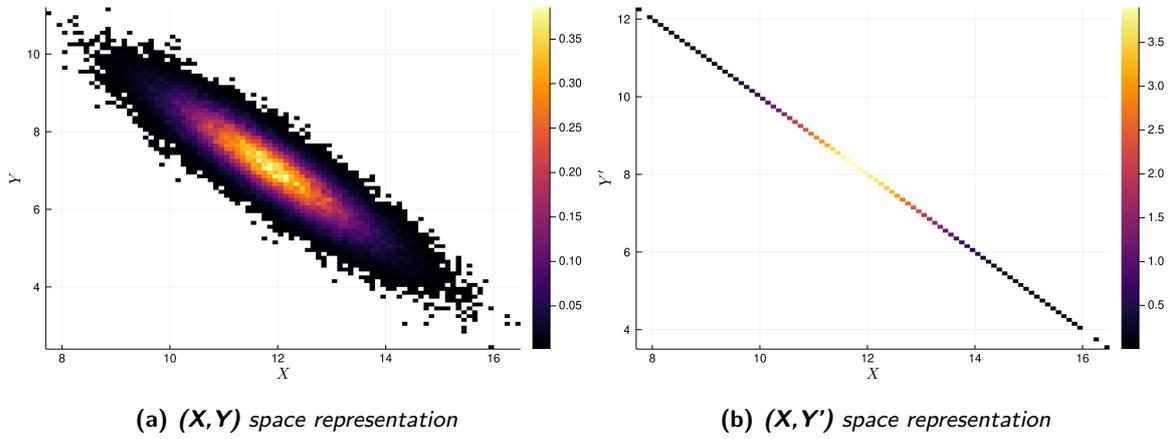
Figure 6.13: *2-dimensional histogram of the stationary probability distribution for an ensemble of 70,000 samples starting from the initial condition $(X_0, Y_0) = (25, 15)$ and simulated until the final time $T=10{,}000$, where the system has reached its stationary profile. The system obeys the dynamical parameters $K = 80$, $A = 0.5$, $\gamma = 0.005$ and $D = 0.0001$.*

To identify the analytical relation of these points on the stability line, the mathematical relation for the straight connecting the steady states $\bar{X}_4$ and $\bar{X}_5$ defined by Eqs. 6.14d and 6.14e in the (X, Y) space was imposed leading to

$$X + Y = 4K_{max} - K. \tag{6.19}$$

Next, another variable change from (X, Y) to (X, Y') was applied with $Y' = 4K_{max} - K - X$ to investigate if the stationary probability distribution is spread along this new relation with the lowest possible variance. The result confirms the general behavior of the system to navigate under noise perturbation along this stability trajectory defined by the deterministic dynamics as shown in Fig.6.13b. The mean of the distribution

corresponds to the final points predicted by the vector field in Fig. 6.5.

6.8.1   Overview of the stationary behavior from computational simulations

From a computational perspective, for each bin on the κ direction, the data was averaged over the corresponding binning in the ζ direction resulting in mean values $\bar{\zeta}$. The result describes the trend of stationary probability distribution in dependence on D (Fig.6.14). As expected, by decreasing D, the samples are progressively more concentrated around the initial position (K_{max}, K_{max}) as a stable point. Consequently, the stationary probability is spread more uniformly by moving away from this point for larger noise strengths. In Fig. 6.14, the resulting behaviors were obtained by using the same number of bins to create the grid on which the histograms were calculated where for the case $D = 0.001$ an increase in the bin length would improve the resolution since the most samples occupy only the last bin. Overall, this analysis enabled to qualitatively study the behavior of the stationary probability distribution towards $\kappa \to K_{max}$, which corresponds with the initial point on the stability manifold in the configuration with $K = 25$ and by changing the noise strength D. Obviously, the noise forces the system to move away from the steady state, which coincides with the initial condition, confirming the role of stochasticity in making the system aware of its alternative steady state configurations.
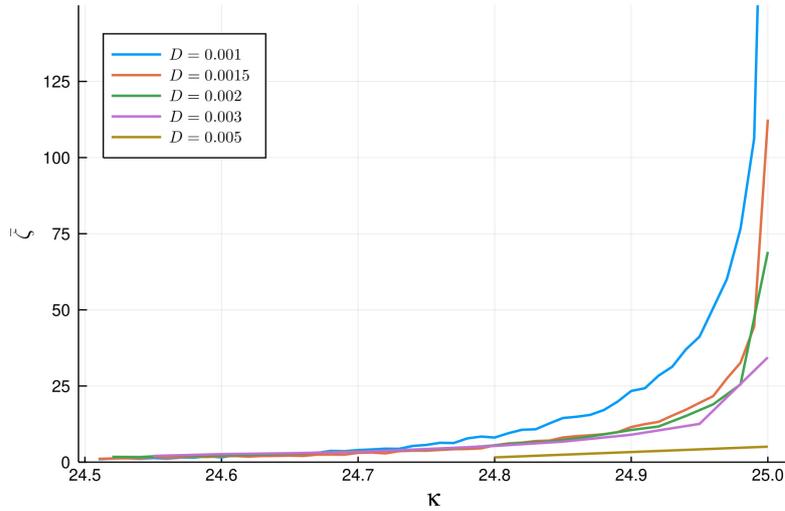
Figure 6.14: *Asymptotic limit of the stationary probability distribution on the manifold when κ is approaching K_{max}. K was fixed at K_{max} and the other parameters are: $A = 0.5$, $\gamma = 0.005$ and (X_0, Y_0)=(25,25). The time simulation is T=8000 and the ensemble size is 40000.*

To further characterize the system's dynamics, I next qualitatively investigated stationary probability

distributions for an initial condition not too close to the critical point K_{max}. For this purpose, computational simulations were performed by varying the noise strength to investigate how these distributions are modified particularly by the underlying stochasticity when the system reaches a steady state. The initial condition was set to $(X_0, Y_0) = (36, 16)$ corresponding to $K = 24$ in the formula for the steady state concentrations on the stability manifold given by Eq. 6.14b. The dynamical parameter K was set to K_{max} and the ensemble size to 40,000 particles representing cells. To allow comparisons between different configurations, the number of bins was fixed for any configuration considered.

(a) *Large bins length*   (b) *Small bins length*
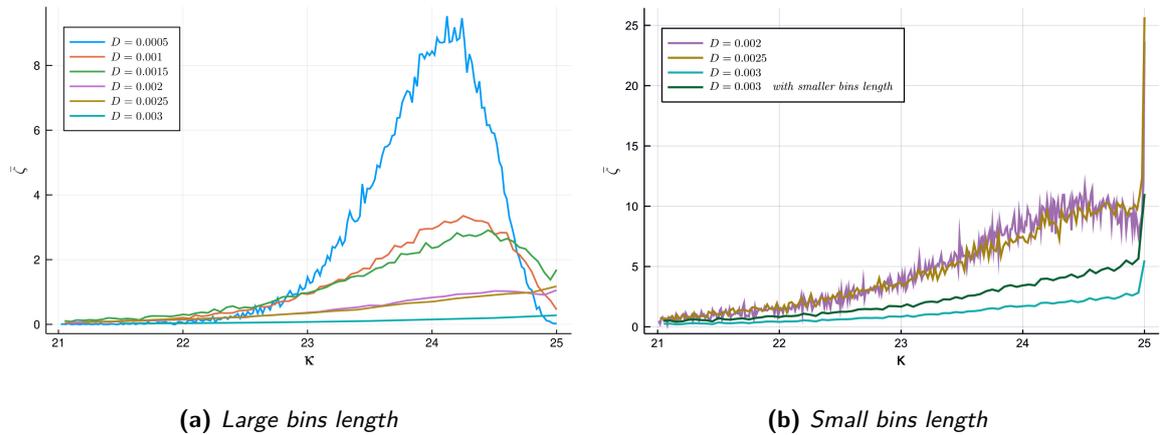
Figure 6.15: *Computational results of the stationary probability distributions on the stability manifold for an statistical ensemble of $M = 40,000$ trajectories starting from $(X_0, Y_0) = (36, 16)$ with $K = K_{max}$ and the usual parameter set.*

The results highlight how the noise affects the shape of the stationary distribution in proximity to the extreme value $\kappa = K_{max}$ (Fig. 6.15). As shown in Fig. 6.15 for $D = 0.001$, the distribution maintains its expected deterministic mean at $\kappa = 24$ with high fluctuations due to the tight binning condition with most of the samples localized in this region. Interestingly, by increasing the noise intensity, the stationary probability distribution gradually looses its preferential steady state in favour of the attractor point at $\kappa = 25$. At $D = 0.003$, the noise strength is able to displace the particles towards a region in the phase space that corresponds to steady state concentrations associated to K similar to K_{max}. This might mean that, when particles by chance move close to K_{max}, they definitively remain trapped within this region of the stationary manifold. Without the driving noise force, this extreme attractive region would remain unexplored.

To verify the hypothesis that noise has an effect in destabilizing the probability distributions when the system is close enough to the extreme point, the stochastic dynamics around another initial position on the

stability trajectory was investigated. For stochastic ensemble simulations with $K = 25$ up to a simulation time of 8000 and with an initial condition defined by $(X_0, Y_0) = (52.3, 7.6)$ that corresponds to a $K = 20$ in accordance to Eq. 6.14b, the results indicate again how noise affects the dynamics of the distributions by exploring the local available space with a referential direction towards K_{max} (Fig. 6.16).

(a) $T=1000$

(b) $T=3000$

(c) $T=8000$
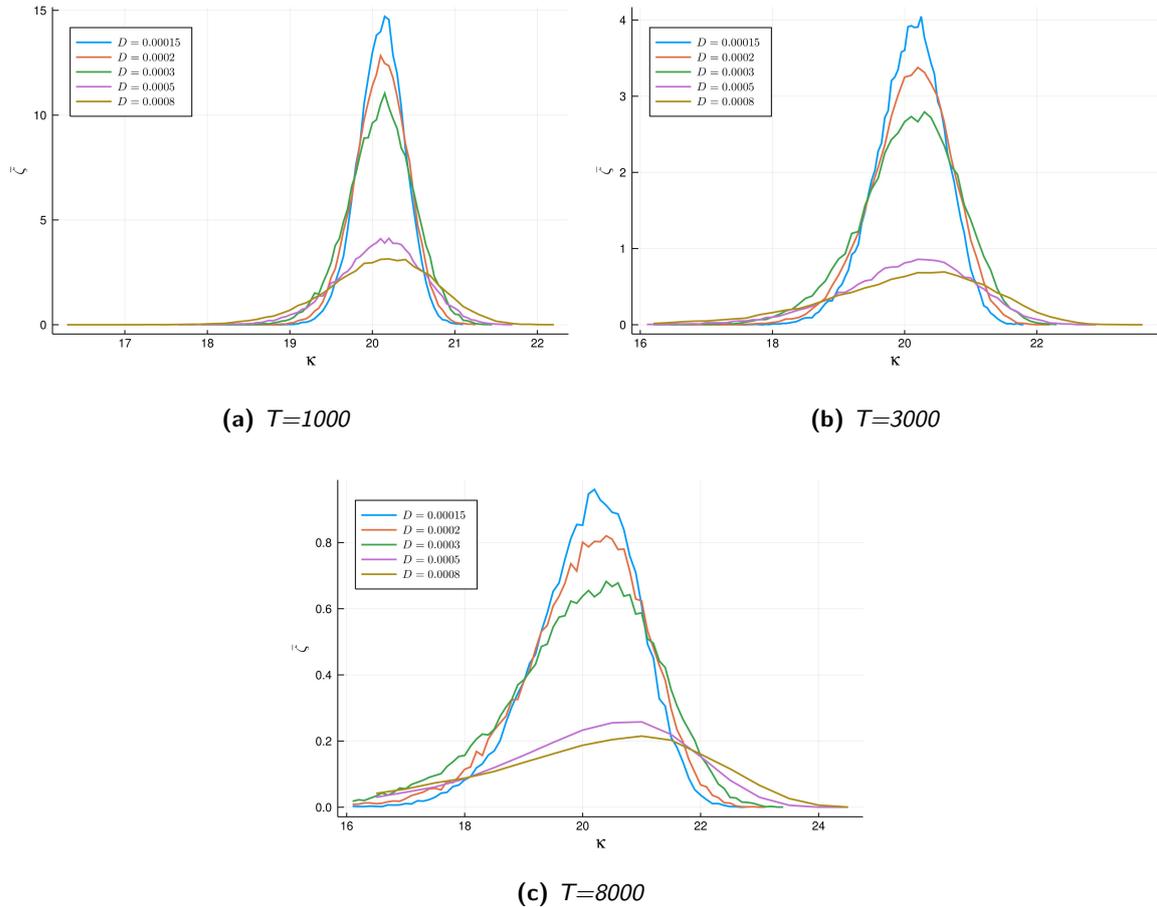
Figure 6.16: *Noise-depending shapes of the stationary probability distribution resulting from an ensemble simulation of 40,000 particles with initial position $(X_0, Y_0) = (52.3, 7.6)$. It is clearly visible how by increasing the noise strength, the particles are located with larger probability far from the initial point and more directed towards $\kappa = 25$.*

As a final analysis of this section, the effect of noise with strength $D = 0.0001$ for an ensemble starting from $(52.3, 7.6)$ was studied as above. I simulated the systems by varying the total time simulation and considering $K = 20$ and $K = 25$. The results are shown in Fig.6.17. The probability distributions change over time until a stationary profile is established around $T = 15,000$ with a specific variance and with the mean corresponding to the the initial steady condition. The stationary localization of the probability

distribution around the initial point was confirmed by longer simulation until the time $T = 25000$. Moreover, the profiles corresponding to $K = 25$ do not show a significant difference compared to those referring to $K = 20$. Taken together, the analyses demonstrated that in the dynamical evolution of the system with a moderate noise level, the starting condition determines the steady state corresponding to the mean of the stationary probability distribution. The choice of K in the deterministic dynamics is irrelevant in determining the stationary profile. Moreover, the noise has thereby the power to spread locally the distributions along the stationary manifold according to the parameter D, demonstrating the adaptability property of the stochastic dynamics.
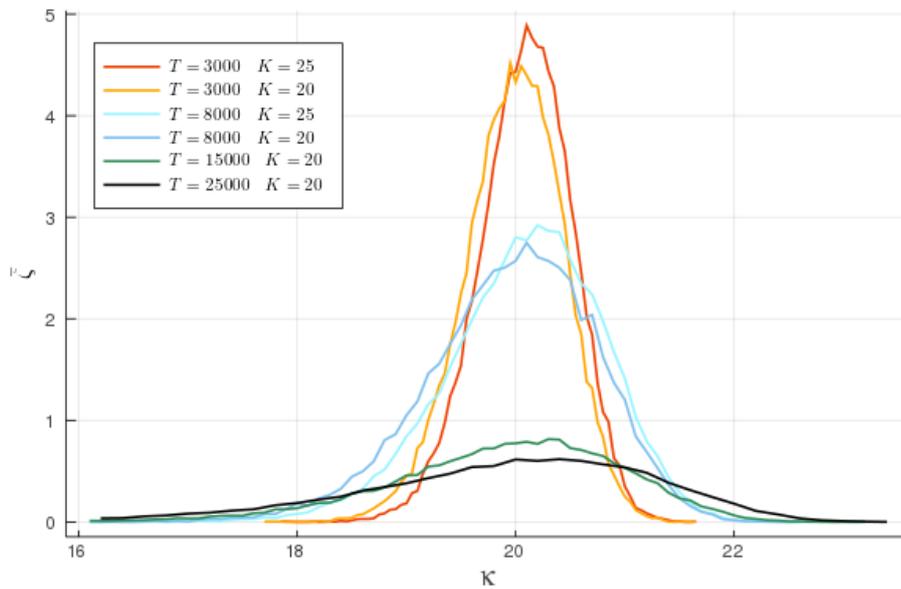
Figure 6.17: *Computational results of the stationary probability distributions over several total times simulation for a stochastic ensemble of 40,000 trajectories starting from (52.3,7.6) on the stability manifold. The noise strength was $D = 0.0001$. The profiles are clearly independent on the parameter K.*

6.8.2 Final remarks on the stationary probability analysis

In this section, computational analysis of the stationary behavior associated to the stochastic one-stage T.S. system in Eq. 6.16 provided insights into the interplay between noise and the deterministic non-linear dynamics. In the regime $K < K_{max}$, the results in Section 6.5 showed that the stability trajectory is traceable by connecting all possible steady states of the dynamics. This was the starting point to understand how the stationary probability distribution are localized on this specific domain in the phase space. The corresponding analysis revealed that stationary concentrations are potentially reachable by the

system in a way strictly dependent on the initial condition and the noise strength.

Indeed, noise allows for tiny displacements over time while the underlying deterministic dynamics occurs. Therefore these random re-locations can conduct the particles (whose states is described by the variable pair (X, Y) to a final state close to the one determined by the deterministic dynamics. Collectively, the continuous noise impact on the dynamics over multiple particles generates probability distributions on the stability manifold. Hence, these results provide mechanistic insights of cell behavior at the first stage of the differentiation process which corresponds to the progenitor destabilization state. The ability to explore the available microstates defined by internal gene regulatory networks is understood by the crucial role of noise that offers alternative local steady states. Moreover, the noise changes significantly the stationary probability distribution for initial condition sufficiently close on the stability manifold. Then the noise induces the system to leave its local stable configuration towards the extreme state with $\kappa = K_{max}$ of the fully symmetry regime. As shown comprehensively in Section 6.8.1, this tendency is linearly dependent on the noise strength that progressively changes the stationary probability distribution from an initially established unimodal, centered distribution to a more decentralized uniform distribution.

6.9   Asymmetric one-stage toggle-switch model

In the previous section, the symmetric case of the toggle-switch model was analysed in detail and revealed how the noisy dynamics can explore the underlying stability manifold. While this symmetric approach identified interesting dynamical behavior which may reflect the induction of cell fate through progenitor state destabilization, the symmetry assumption might only hold in specific biological circumstances or limited time periods. To investigate how the more general case of an asymmetric relation between the genes X and Y affect the dynamics, this section is focusing on the stage of the cellular differentiation process in which an asymmetry is locally well-established in the metastable state arising from the progenitor state destabilization when the cell is driven towards its closest stable state. For this process, I assume that the resulting stochastic dynamics evolves according to the poised internal gene configuration and the asymmetry of the underlying deterministic dynamics which can be imposed by considering different value for K_X and K_Y in the one-stage model. With this assumptions, the deterministic dynamics takes the form

$$\begin{cases} \dot{X} = A \frac{K_Y}{K_Y+Y} \frac{X}{K_X+X} - \gamma X \\ \dot{Y} = A \frac{K_X}{K_X+X} \frac{Y}{K_Y+Y} - \gamma Y \end{cases} \tag{6.20}$$

with its representative scheme shown in Fig. 6.18.
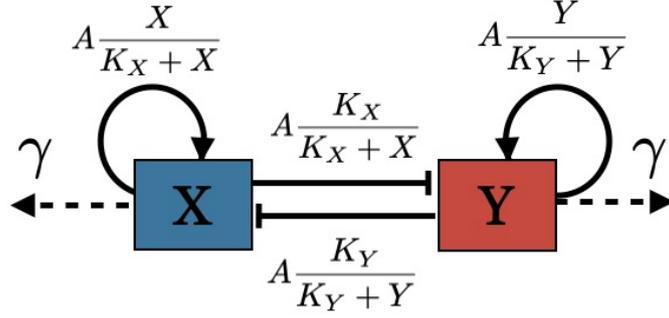
Figure 6.18: *Asymmetric interactions in the one-stage T.S. model where K_X and K_Y regulate the strength of self-activation and cross-inhibition terms.*

The corresponding stability analysis allows for the identification of the following three fixed points: $\bar{X}_0 = (0,0)$, $\bar{X}_1 = \left(\frac{A}{\gamma} - K_X, 0\right)$ and $\bar{X}_2 = \left(0, \frac{A}{\gamma} - K_Y\right)$. To investigate the nature of these states, the Jacobian matrix of the system and its eigenvalues for each point is evaluated in analogy to Section 6.6. The Jacobian matrix reads

$$J(\{A, \gamma, K_X, K_Y\}) = \begin{pmatrix} \frac{AK_XK_Y}{(K_Y+Y)(K_X+X)^2} - \gamma & \frac{-AK_YX}{(K_X+X)(K_Y+Y)^2} \\ \frac{-AK_XY}{(K_Y+Y)(K_X+X)^2} & \frac{AK_XK_Y}{(K_X+X)(K_Y+Y)^2} - \gamma \end{pmatrix}$$

where the inequality $\frac{A}{\gamma} \geq K_X, K_Y$ must hold for positive concentrations. The fixed point $\bar{X}_0$ is an unstable state for any feasible positive parameter setting. The Jacobian matrix evaluated at $\bar{X}_1$ can be written as

$$J(\{A, \gamma, K_X, K_Y\})|_{\bar{X}_1} = \begin{pmatrix} \gamma\psi_X & \frac{A}{K_Y}\psi_X \\ 0 & \gamma\pi_X \end{pmatrix}$$

with $\psi_X \equiv \frac{K_X}{K_{max}} - 1$, $\pi_X \equiv \frac{K_X}{K_Y} - 1$ and $K_{max} = \frac{A}{\gamma}$. With these identifications, the eigenvalues are given by $\lambda_1 = \gamma\psi_X$, which is always negative since $K_X \leq K_{max}$ and $\lambda_1 = \gamma\pi_X$ whose sign depends on K_X and K_Y. If $K_X \leq K_Y$, $\bar{X}_1$ is stable otherwise it is unstable. Analogously, for the fixed point $\bar{X}_2$ the corresponding Jacobian matrix is given by

$$J(\{A, \gamma, K_X, K_Y\})|_{\bar{X}_2} = \begin{pmatrix} \gamma\pi_Y & 0 \\ \frac{A}{K_X}\psi_Y & \gamma\psi_Y \end{pmatrix}$$

where $\psi_Y \equiv \frac{K_Y}{K_{max}} - 1$ and $\pi_Y \equiv \frac{K_Y}{K_X} - 1$. Hence, if $K_Y < K_X$, $\bar{X}_2$ is a stable state and $\bar{X}_1$ is unstable otherwise $\bar{X}_2$ is the unstable state and $\bar{X}_1$ is the stable one. In the case $K_X = K_Y$, one eigenvalue is zero and the other is negative. In analogy to the symmetric case in Section 6.6, the vector field is visualized by the tangent vector at each point on a grid. In this way, the instantaneous tendency of the dynamics for any initial condition fixed on the grid can be observed. The results confirm the analytic derivation of the steady

states as shown in Fig.6.20. The graphs contain the dashed line describing the straight line through $\bar{X}_1$ and $\bar{X}_2$ given by

$$X(K_{max} - K_Y) + Y(K_{max} - K_X) - (K_{max} - K_X)(K_{max} - K_Y) = 0. \qquad (6.21)$$

For $K_X = K_Y$, the relation between the steady states $\bar{X}_4$ and $\bar{X}_5$ of the symmetric case is recovered (see Eq. 6.19) as shown in Fig. 6.5. Thereby the cells might be subjected to a structural change of the phase space due to an underlying evolutionary dynamics from a symmetric through a slight asymmetry configuration that changes abruptly the nature of the steady states.
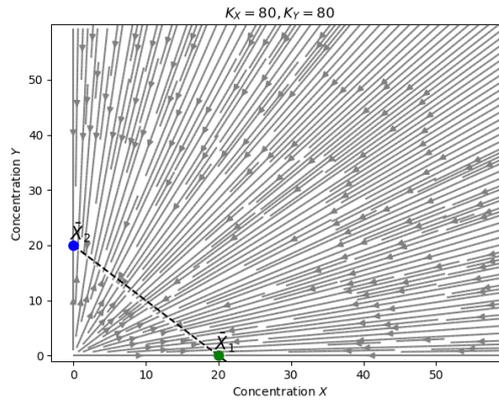
Figure 6.19: *The vector field for $K_X = K_Y = 80$ is the same obtained for the symmetric case in Fig. 6.5 as expected.*

(a) $K_X > K_Y$ **(b)** $K_Y > K_X$
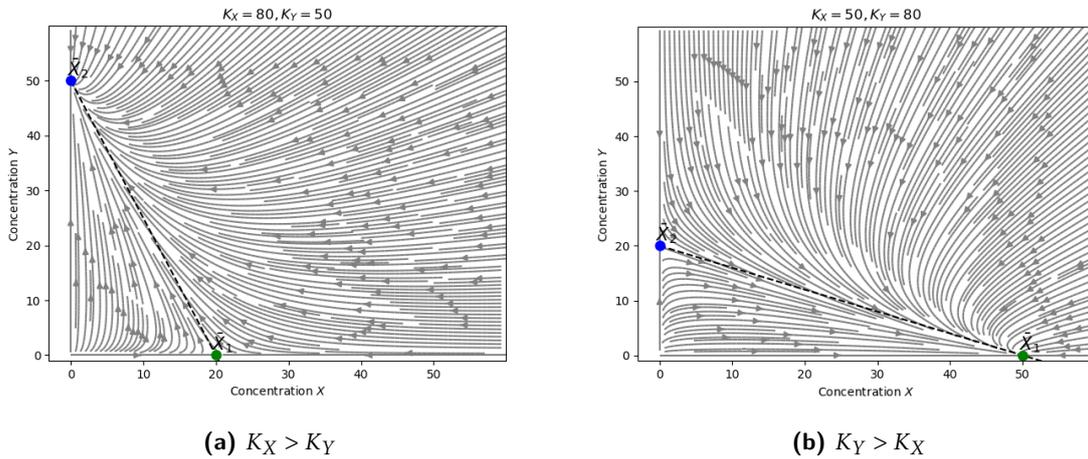
Figure 6.20: *The vector field of the asymmetric one-stage T.S. model underline the existence of one stable state and one unstable state. The dashed lines correspond to the straight line through $\bar{X}_1$ and $\bar{X}_2$.*

By generating an asymmetry in the interaction strengths, the nature of the steady states changes and consequently the underlying dynamics. The vector fields show a preferential direction which connects the stable to the unstable state equivalently to the symmetric case in the regime $K \geq K_{max}$. Hence, the system is pushed faster towards the stable state on this trajectory.

6.9.1 Stochastic analysis of the asymmetric model in the Fokker-Planck formulation

With the analogous goal as in the previous FPE analysis, I studied the bivariate Langevin Equations with white noise for the asymmetric case. In this formulation, the noise might contribute to an addiction driving force in the system by accelerating or restraining its deterministic dynamics over the phase space. The SDEs in the asymmetric reads

$$\begin{cases} \dot{X} = \frac{AK_Y}{K_Y+Y} \frac{X}{K_X+X} - \gamma X + \eta_X(t) \\ \dot{Y} = \frac{AK_X}{K_X+X} \frac{Y}{K_Y+Y} - \gamma Y + \eta_Y(t) \,, \end{cases} \tag{6.22}$$

with $\eta_X(t)$ and $\eta_Y(t)$ being random forces with the same statistical properties as in Eq. 6.17.
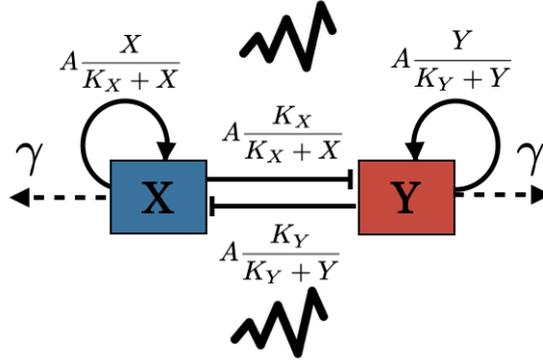
Figure 6.21: *Asymmetric interactions in the noisy one-stage T.S. model. K_X and K_Y regulate the strength of self-activation and cross-inhibition terms. The dynamics is perturbed by white noise at each time step.*

Here, I exploited computational simulations to qualitatively investigate Eqs. 6.22 over time and by varying the diffusion coefficient D, which was put equal for X and Y for simplicity.

First, I studied the configuration with the initial condition $(X_0, Y_0) = (0.1, 18)$ and $D = 0001$ for an ensemble of $M = 40,000$ particles and a total simulation time T of 4000 (Fig. 6.20a). To visualize the

dynamics, 2-dimensional histogram of all X and Y concentrations at time points $T = 250$, $T = 750$, $T = 1000$, $T = 1500$, $T = 2500$ and $T = 4000$ where generated (Fig. 6.22a). For this configuration with an initial condition close to the unstable state, the system approaches first this unstable state and then follows the straight line (corresponding to dashed line in Fig. 6.20a) to finally reach the stable state $\bar{X}_2$. Given the analytical expression of that trajectory in Eq. 6.21, Y can be expressed by means of X and vice versa leading to

$$Y = K_{max} - K_Y - X\left(\frac{K_{max} - K_Y}{K_{max} - K_X}\right) \equiv Y', \tag{6.23}$$

$$X = K_{max} - K_X - Y\left(\frac{K_{max} - K_X}{K_{max} - K_Y}\right) \equiv X'. \tag{6.24}$$

By heuristically finding this domain of the probability distributions on the phase space (Figs. 6.20a and 6.20b), this line can reduce the degrees of freedom of the stochastic dynamics to one and allow to study the variability of X and Y as independent variables. Eq. 6.23 can be rewritten as Y' and Eq. 6.24 as X' to highlight that Y' (X') expresses the variability of X (Y) restricted on a domain defined by the relationship with Y (X). These dependencies can be used to separately represent the X and Y concentration probability distributions over the trajectory in Eq. 6.21 during the temporal evolution as shown in Fig. 6.22.
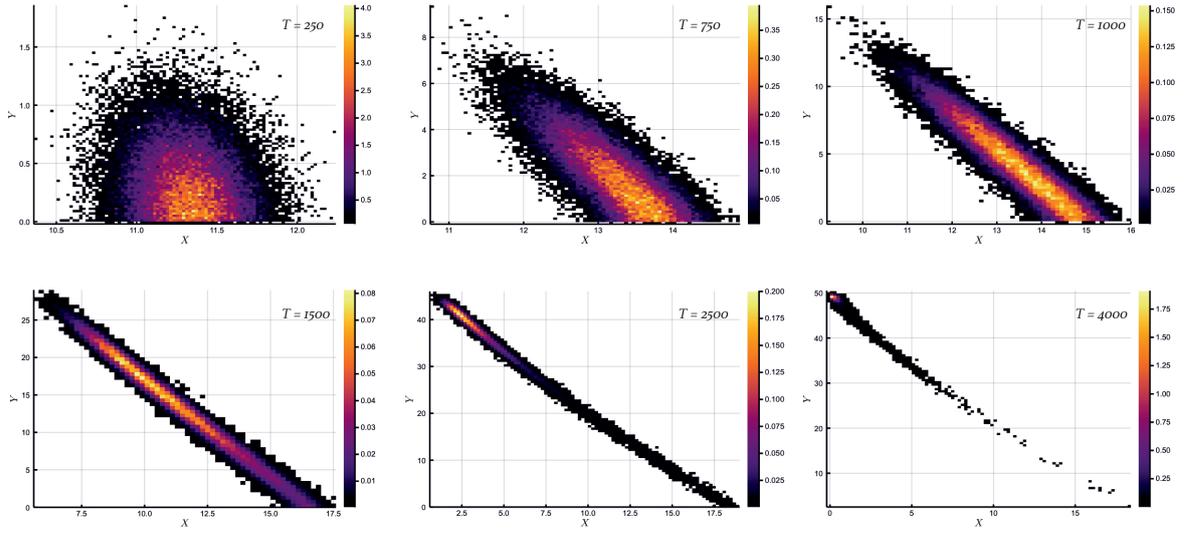
(a) *Temporal evolution of the 2-dimensional probability distribution on the phase space.*
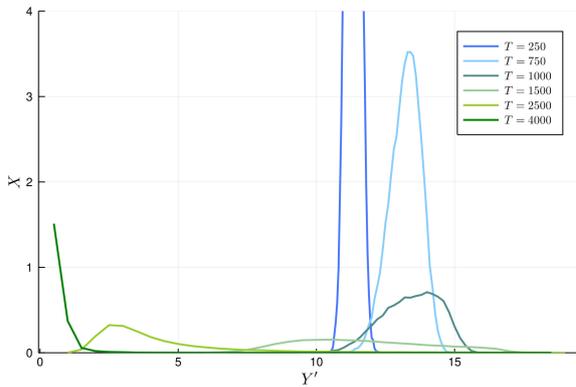
(b) *X-component probability distribution*
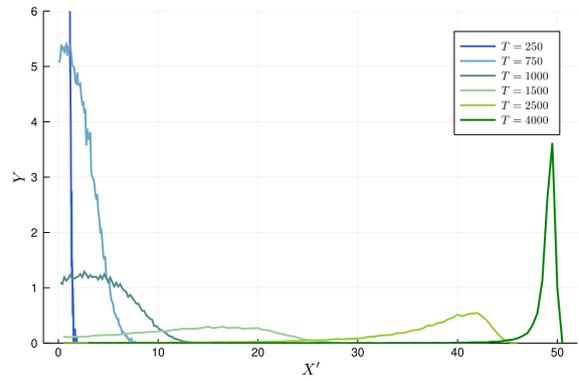
(c) *Y-component probability distribution*

Figure 6.22: *Approximation of temporal evolution of the joint probability distribution as an 2-dimensional histogram on the phase space in Fig. 6.22a and the corresponding marginal probability for X (in Fig. 6.22b) and Y (in Fig. 6.22c) on the straight line connecting $\bar{X}_1$ and $\bar{X}_2$. The initial condition is $(X_0, Y_0) = (10, 0.1)$ for an ensemble of 40,000 cells and dynamics parameters $A = 0.5$, $\gamma = 0.005$, $K_X = 80$, $K_Y = 50$ and $D = 0.0001$.*

Since the noise strength was set to a rather small value $(D = 0.0001)$, the probability distribution follows the deterministic dynamics imposed by the initial condition since the mean is mainly moving along the corresponding vector flux and specifically on the line defined in Eq. 6.21. Therefore, the noise affection on the deterministic dynamics emphasizes the intrinsic variability of the system generated by the continuous interplay with a stochastic source. Next, the same system was investigated with a stronger diffusion coefficient of $D = 0.1$ (Fig. 6.23). As expected, the distribution evolves faster compared to the previous case but the corresponding variances are also larger. Moreover, the noise pushes the system to follow a

direction faster towards the stable state since it can jump to other lines of the vector field without crossing the unstable state.
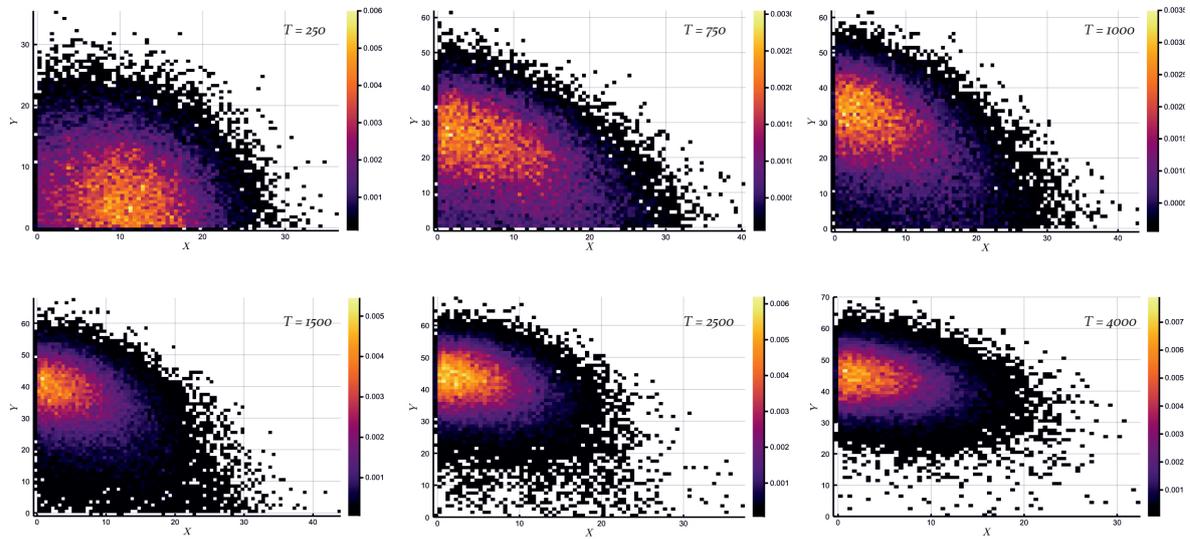
(a) *Temporal evolution of the 2-dimensional probability distribution on the phase space.*

(b) X-*component probability distribution*

(c) Y-*component probability distribution*
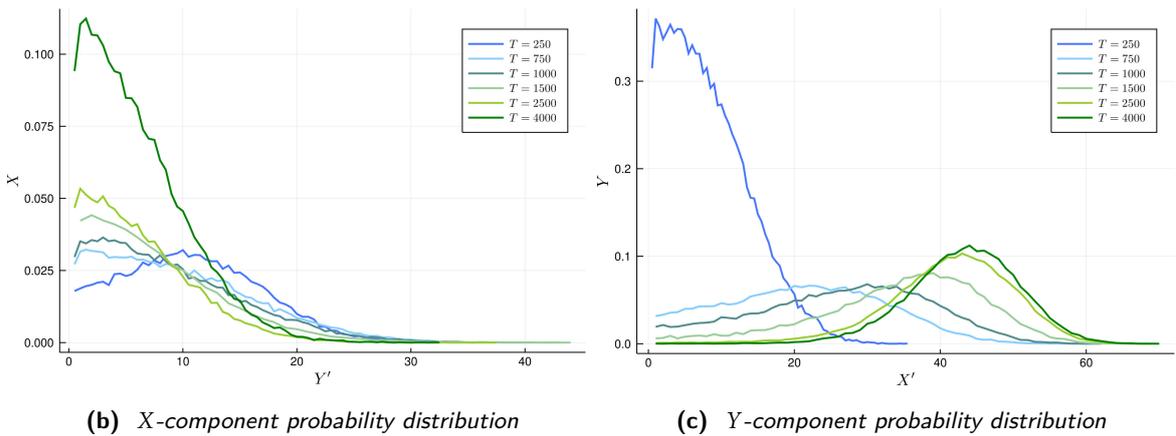
Figure 6.23: *Approximation of temporal evolution of the joint probability distribution as an 2-dimensional histogram on the phase space in Fig. 6.23a and the corresponding marginal probability for X (in Fig. 6.23b) and Y (in Fig. 6.23c) on the straight line in Eq. 6.21 connecting $\bar{X}_1$ and $\bar{X}_2$. The initial condition is $(X_0, Y_0) = (10, 0.1)$ for an ensemble of 40,000 cells and dynamics parameters $A = 0.5$, $\gamma = 0.005$, $K_X = 80$, $K_Y = 50$ and $D = 0.1$.*

In Fig. 6.24, the dynamics of the deterministic molecular concentrations are plotted to appreciate the noise-induced acceleration in pushing the system faster to the stable state.
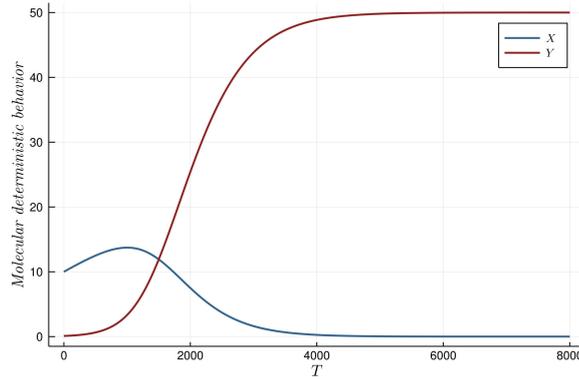
Figure 6.24: *Deterministic behavior of the asymmetric one-stage T.S. with the same parameters used above:* $A = 0.5$, $\gamma = 0.005$, $K_X = 80$, $K_Y = 50$ *with initial condition* $(X_0, Y_0) = (10, 0.1)$.

These computational evidences of the asymmetric toggle-switch demonstrate how the induction of cell fate commitment by progenitor destabilization and subsequent is affected by noise. In particular, initial conditions closer to the unstable steady state have the tendency to first approach this unstable state before eventually ending up in the stable state in the case when noise is not so strong to deviate the system respect the deterministic trajectory as shown in Fig. 6.22. From a population perspective, this mechanism might be an explanation for the observation of "rebellious" cells [39], where progenitor cells are first destabilized and randomly distributed at low X and Y concentrations. In dependence on the induced stable steady states, cells with an initial condition closer to the unstable state will first approach the unstable attractor before eventually move towards the stable state along the connecting line.

6.10   Symmetric two-stage toggle switch model

As mentioned in the Introduction and Chapter 3, cell fate and heterogeneity originate form the hierarchical biological organization across different layers Fig. 3.4. In this section, I present a more comprehensive version of the symmetric toggle-switch model characterized by a four-dimensional ODEs system which reflects the biological levels of transcription and translation and their interplay in line with the central dogma of molecular biology [166] . In this respect, the approach is similar to previous work [33] but uses an alternative methodology as detailed in Section 6.5. Basically, the stochasticity is here introduced as a continuous modification of molecular concentrations in a mesoscopic description and not considered as intrinsic fluctuations of low copy numbers of molecules for which the Master equation [167] of the system would correspond to the probabilistic associated description [33].

The mRNAs production is based on the assumption that DNA is always available for the transcription.

Its change in concentration evolves by following a mechanism of regulation which involves both protein concentrations in a term defined by the product between non-linear cross-inhibition and self-activation factors. Specifically, the model I proposed is schematically and mathematically described in Figure 6.25 and as follows:

$$
\begin{cases}
\dot{X} = \alpha x - \gamma X \\
\dot{Y} = \alpha y - \gamma Y \\
\dot{x} = A\frac{X}{K+X}\frac{K}{K+Y} - K^d x \\
\dot{y} = A\frac{Y}{K+Y}\frac{K}{K+X} - K^d y
\end{cases}
\tag{6.25}
$$
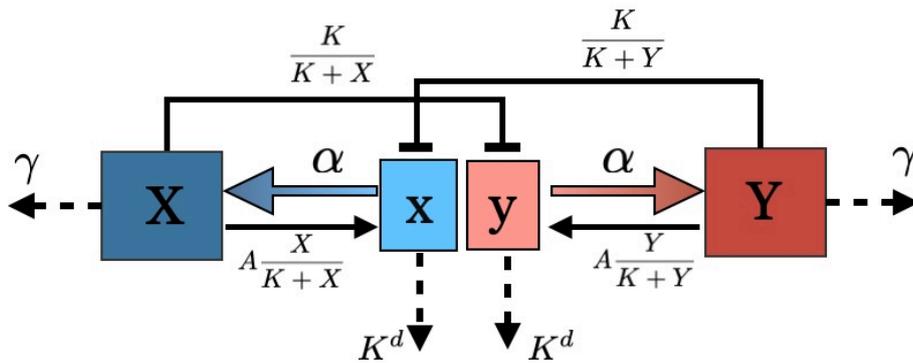
Figure 6.25: *Scheme of the two-stage symmetric toggle switch model in which the transcription and translation dynamics are distinguished into two interacting regulatory levels. The protein production rate is linearly dependent on the associated mRNA and itself-concentrations. Whereas mRNAs production is regulated by the product of non-linear terms representing self-activation and cross-inhibition acted by the protein concentrations.*

The splitting into the two interacting regulatory mechanisms is firstly motivated by the biological mechanism that cell dynamics is internally tuned by the interactions between the genotype and phenotype by different layers of regulation (3.4,[27]). Secondly, the aim is to investigate the role of an additional level of regulation under a noisy faster time-scale dynamics in stabilizing and maintaining a steady state configuration for the whole system. Since the protein production is a slower [168] process compared to the mRNA transcription, it can operate as a buffer by reducing noise effects of the more stochastic mRNA dynamics. To investigate general mechanisms underlying cell fate commitment during the differentiation process, this two-stage model aims at discovering how the final phenotype choice is triggered at the protein

level in presence of gene expression variability.

In agreement with the hypothesis that the cell explores its own local available space by assuming intermediate possible configurations, the choice of symmetric interaction between the competitive cell fates is an appropriate assumption to describe this transitory state based on the previous analysis of the one-stage toggle switch model above. In order to evaluate analytically the degenerate steady states of a corresponding dynamical system, I start from a generic two-stage dynamics in which the evolution of the mRNA dynamics denoted by x and y, respectively, is modulated by the same parameter for self-activation and cross-inhibition through K_X and K_Y. While in the one-stage model (Sections 6.5-6.6) these interactions were more globally defined, the two-stage model allows for a more biological interpretation where the mRNA dynamics is regulated by the protein or transcription factor level. A generic ODEs system of the two-stage model can be written as

$$
\begin{cases}
\dot{X} = \alpha x - \gamma X \\
\dot{Y} = \alpha y - \gamma Y \\
\dot{x} = A \frac{X}{K_X + X} \frac{K_X}{K_X + Y} - K^d x \\
\dot{y} = A \frac{Y}{K_Y + Y} \frac{K_Y}{K_Y + X} - K^d y \,.
\end{cases}
\tag{6.26}
$$

By substituting the first two linear relationships at equilibrium $x = \frac{\gamma}{\alpha} X$ and $y = \frac{\gamma}{\alpha} Y$ into the mRNA dynamics of x and y, the trivial equilibrium state $(0,0)$ and the two extreme cases $\bar{X}_4 = \left(0, 0, \frac{A\alpha}{K^d \gamma} - K_Y, \frac{A}{K^d} - \frac{\gamma K_Y}{\alpha} \right)$ corresponding to a zero concentration for X and $\bar{X}_5 = \left(\frac{A\alpha}{K^d \gamma} - K_X, \frac{A}{K^d} - \frac{\gamma K_X}{\alpha}, 0, 0 \right)$ corresponding to a zero concentration for Y can be identified. The other fix points are obtained by imposing non-vanishing concentrations for both molecules and by following the same calculations performed in the reduced one-stage model in Section 6.6 leading to similar equilibrium concentrations as in the one-stage dynamics. The slight difference is in the parameter rearrangement of $Y_{max} \equiv \frac{A\alpha}{\gamma K^d} - K_X - K_Y$ expression which was found from the relationship $X^{(0)} + Y^{(0)} = K_{max}$ and leading to the same two steady-state concentrations as in one-stage model given by

$$
Y_{\pm}^{(0)} = X_{\pm}^{(0)} = \frac{Y_{max}}{2} \pm \frac{\sqrt{Y_{max}^2 - 4 K_X K_Y}}{2}.
\tag{6.27}
$$

The steady states can then be rearranged by

$$
\begin{cases}
\bar{X}_2 \equiv \left(Y_+^{(0)}, y_+^{(0)}, Y_-^{(0)}, y_-^{(0)} \right) \\
\bar{X}_3 \equiv \left(Y_-^{(0)}, y_-^{(0)}, Y_+^{(0)}, y_+^{(0)} \right),
\end{cases}
\tag{6.28}
$$

where $y_\pm^{(0)} = \frac{\gamma}{\alpha} Y_\pm^0$. The explicit solution for the symmetric case can be recovered by putting $K_X = K_Y = K$ leading to

$$Y_\pm^{(0)} = \frac{1}{2}\left(-2K + \frac{A\alpha}{K^d\gamma}\right) \pm \frac{1}{2}\sqrt{\left(\frac{A\alpha}{K^d\gamma}\right)^2 - \frac{4A\alpha K}{K^d\gamma}}. \tag{6.29}$$

By imposing the existence of physical meaningful solutions in the steady states $\bar{X}_2$ and $\bar{X}_3$, an upper limit for K is given by $K \le \frac{A\alpha}{4K^d\gamma} \equiv K_{max}$. A second extreme value related to the positive concentrations of molecules in the other two states $\bar{X}_4$ and $\bar{X}_5$ given by $\frac{A\alpha}{K^d\gamma}$.

In analogy to Section 6.6, the nature of the steady states in dependence on K as a bifurcation parameter was assessed by a corresponding *Python* script. The analysis shows that the eigenvalue of the Jacobian matrix evaluated at the steady points with the biggest real part is zero. To characterize the dynamics in the phase space, the local dynamics can be considered in a quasi steady-state approximation exploiting the different time scale of transcription and translation by assuming that the faster mRNA dynamics reaches immediately a steady states associated to instantaneous protein concentrations under the dynamics of X and Y. With this assumption, the ODEs system is describing the evolution of the protein concentrations under a quasi-steady state assumption for the mRNAs production reduced to

$$\begin{cases} \dot{X} = \frac{A\alpha K}{K^d} \frac{X}{(K+X)(K+Y)} - \gamma X \\ \dot{Y} = \frac{A\alpha K}{K^d} \frac{Y}{(K+X)(K+Y)} - \gamma Y. \end{cases} \tag{6.30}$$

With this approximation, the vector field in the regions of the phase space around $\bar{X}_2$ and $\bar{X}_3$ can be visualized in 2 dimensions as show in Fig. 6.26 with parameters $A = \alpha = 0.05$, $K^d = 0.005$, $\gamma = 0.0005$ for configuration $K_{max} = 250$. A comparison with the configuration of $K = 220$ shown in Fig. 6.26 shows that the behavior of the system around the points is the same reflecting the symmetry of the system in X and Y. Te points $\bar{X}_2$ and $\bar{X}_3$ in the plots are the coordinates found by the analytic computation.
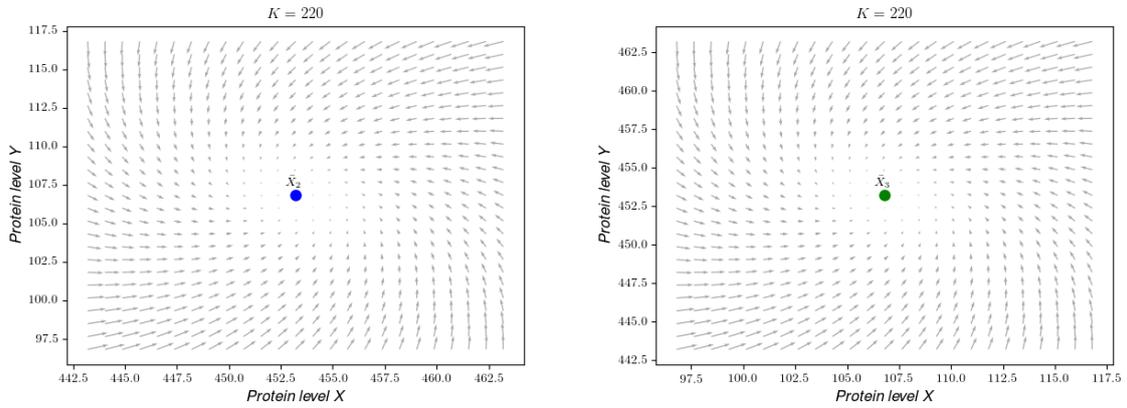
Figure 6.26: *Local vector field of the protein dynamics in the quasi-steady state approximation around $\bar{X}_2$ (on the left) and $\bar{X}_3$ (on the right) in the case $K = 220$.*

Next, I investigated the instantaneous changes of the system with K close to the extreme value $K_{max} = 250$ (Fig.6.27).
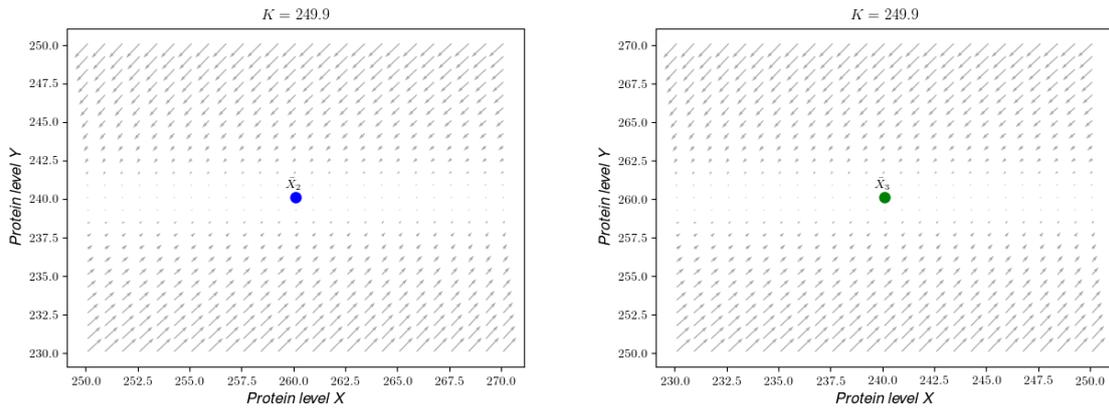
Figure 6.27: *Local vector field of the protein dynamics in the quasi-steady state approximation around $\bar{X}_2$ (on the left) and $\bar{X}_3$ (on the right) in the case $K = 249.9$.*

Interestingly, the system behavior around the steady states $\bar{X}_2$ and $\bar{X}_3$ changes by varying K. I computationally evaluated the signs of the real part of the eigenvalues for the Jacobian matrix at these points and observed that one eigenvalue is zero and one is negative for any $K \leq K_{max}$. In order to confirm these trends I simulated the dynamics over a time window starting from different initial conditions (Fig.6.28) indicating the presence of a line of steady states. It seems that by increasing the vicinity of the two possible steady states, meaning for K approaching K_{max}, the vector flux of the system highlights its tendency to approach rapidly a steady state configuration by changing the curved slow trend to a more straightforward

98

direction to the multiple steady states on a vertical line in the phase space. This is in agreement with the findings for the fully symmetric one-stage toggle switch in Section 6.6 since the system is in a similar fully symmetric configuration.
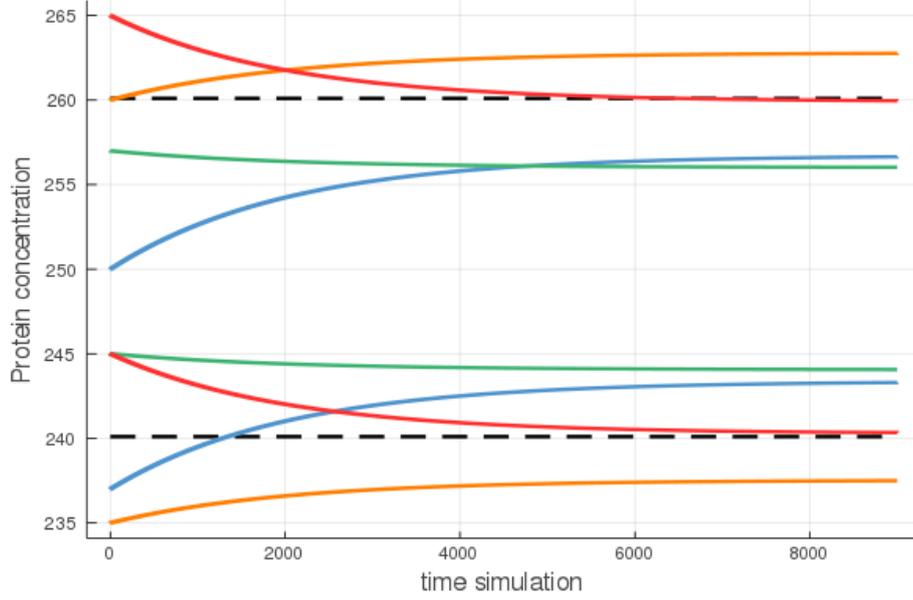
Figure 6.28: *Protein concentration trends depending on the initial conditions for $K = 249.9$.*

In the vicinity of $\bar{X}_2$ and $\bar{X}_3$ the manifold of the system seems to be modified allowing to create phenotype variability on a stability line. Therefore the system seems to be able to adjust its steady state configurations when two regions of stability approach each other. To investigate the effect of noise on the local deformation, a Gaussian white noise was added to the transcriptional level for both variables leading to the corresponding stochastic differential equations of the system

$$
\begin{cases}
\dot{X} = \alpha x - \gamma X \\[4pt]
\dot{Y} = \alpha y - \gamma Y \\[4pt]
\dot{x} = A \frac{X}{K_X + X} \frac{K_X}{K_X + Y} - K^d x + \eta_X(t) \\[4pt]
\dot{y} = A \frac{Y}{K_Y + Y} \frac{K_Y}{K_Y + X} - K^d y + \eta_Y(t) \, ,
\end{cases}
\tag{6.31}
$$

with $\eta_X(t)$ and $\eta_Y(t)$ being random forces with the same statistical properties as in Eq. 6.17. In order to

99

investigate the noisy interfere between the two levels of regulation, a stochastic ensemble was simulated and compared with the noisy one-stage toggle switch. As a first configuration, the parameters were chosen as $A = 0.5, \alpha = 0.05, \gamma = 0.0005, K^d = 0.005$ and $K = 2500$ for the two-stage system and $A = 0.5$, $\gamma = 0.005$ and $K = 25$ for the one-stage system. For both systems, the noise strength was set to $D = 0.01$ and an ensemble of $M = 40,000$ cells were simulated for a simulation time of $T = 1000$. The initial conditions for the corresponding two-stage system were set to $X_0 = 2650$, $Y_0 = 2450$, $x_0 = 26.5$ and $y_0 = 24.5$ and for the one-stage to $X_0 = 26.5$ and $Y_0 = 24.5$. These initial conditions are close to the stable states for both systems and the deterministic dynamics is able to reach the steady point within the time simulation for both systems. The aim of this investigation is to characterize how the phenotype distributions evolve over time and quantify differences induced by the noisy dynamics between the one-stage and two stage toggle switch models. For this purpose, the distributions were overlapped by dividing the concentrations respect to their corresponding average values to allow for quantitative comparisons. The results show how the variability on the protein level is decreased compared to the mRNA level and lead to a stabilization of the corresponding phenotype Fig. 6.29. This indicates that the feedback loop dynamics at different time scales is necessary for the system to moderate the effect of gene variability.
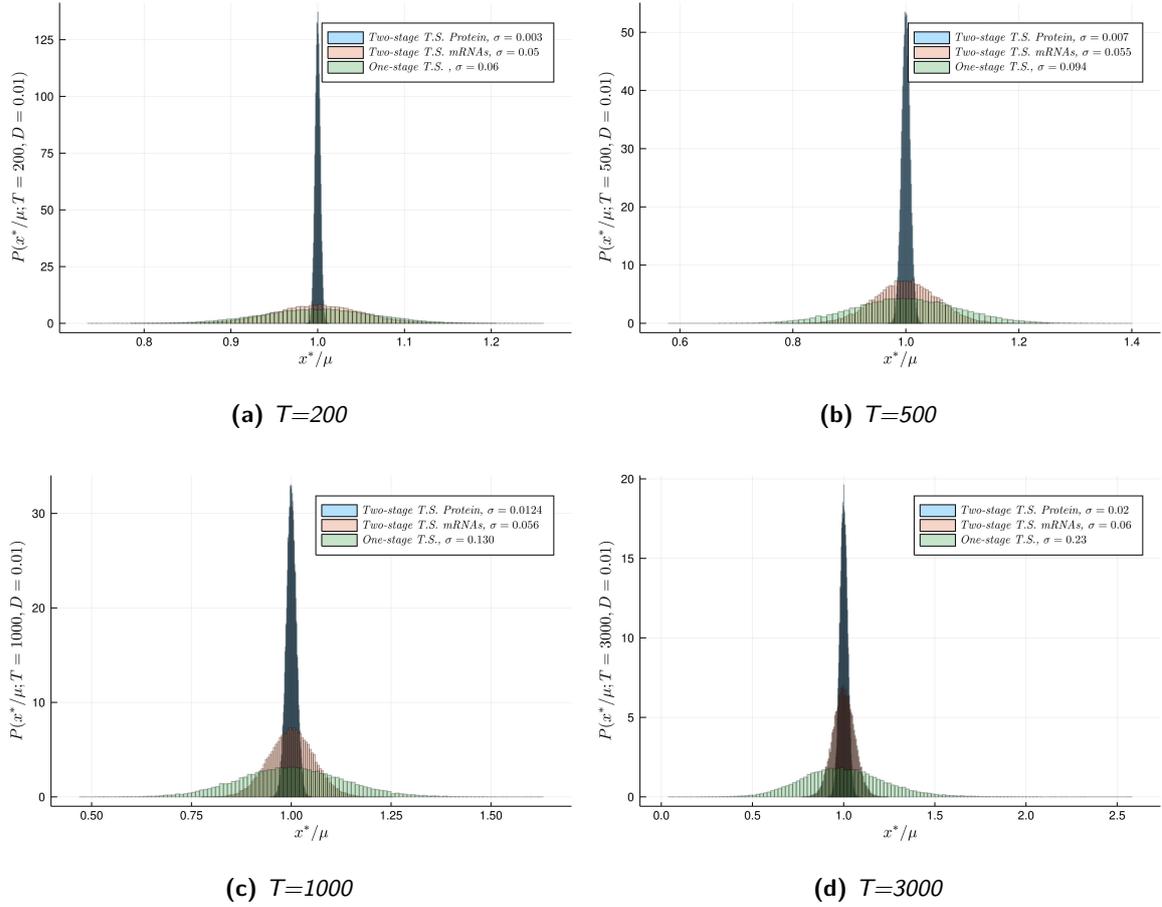
Figure 6.29: *Time evolution of the phenotype distributions of the 1-dimensional random variable x^*/μ referred to the protein concentration in two-stage toggle switch model (in cyan), for the mRNAs concentration at two-stage model (in light pink) and the lumped level of both (in light green). The parameter setting is given by $A = 0.5, \alpha = 0.05, \gamma = 0.0005, K^d = 0.005$ and $K = 2500$ for the two-stage system and $A = 0.5$, $\gamma = 0.005$ and $K = 25$ for the one-stage system. For both system the noise strength was set to $D = 0.01$, the ensemble size to $M = 40,000$ and simulation time to $T = 1000$. The initial condition is sufficiently close to the steady state.*

To demonstrate the time scale separation as a requirement for the buffering effect of the protein level, the phenotype distributions were next investigated for the case where the basal transcription and translation rate are identical. Indeed, this seems to contradict the biological need of slowing down the dynamics to maintain adaptability and gene-variability without loosing robustness and directionality. For these simulations, $A = \alpha = 0.05$ was used and the other parameters are unchanged. The initial condition was adapted consequently to maintain the vicinity to the steady state. The resulting distribution highlight how the protein distribution of the two-stage system progressively sense the downstream noisy level of transcription and increase its intrinsic variability without the time scale separation (in Fig. 6.30). At the

steady state it exactly overlaps with the stationary phenotype distribution of the one-stage models whereas the mRNAs distributions look like Gaussian noise with a variance that increases over time.

(a) *T=200*

(b) *T=500*

(c) *T=1000*

(d) *T=3000*
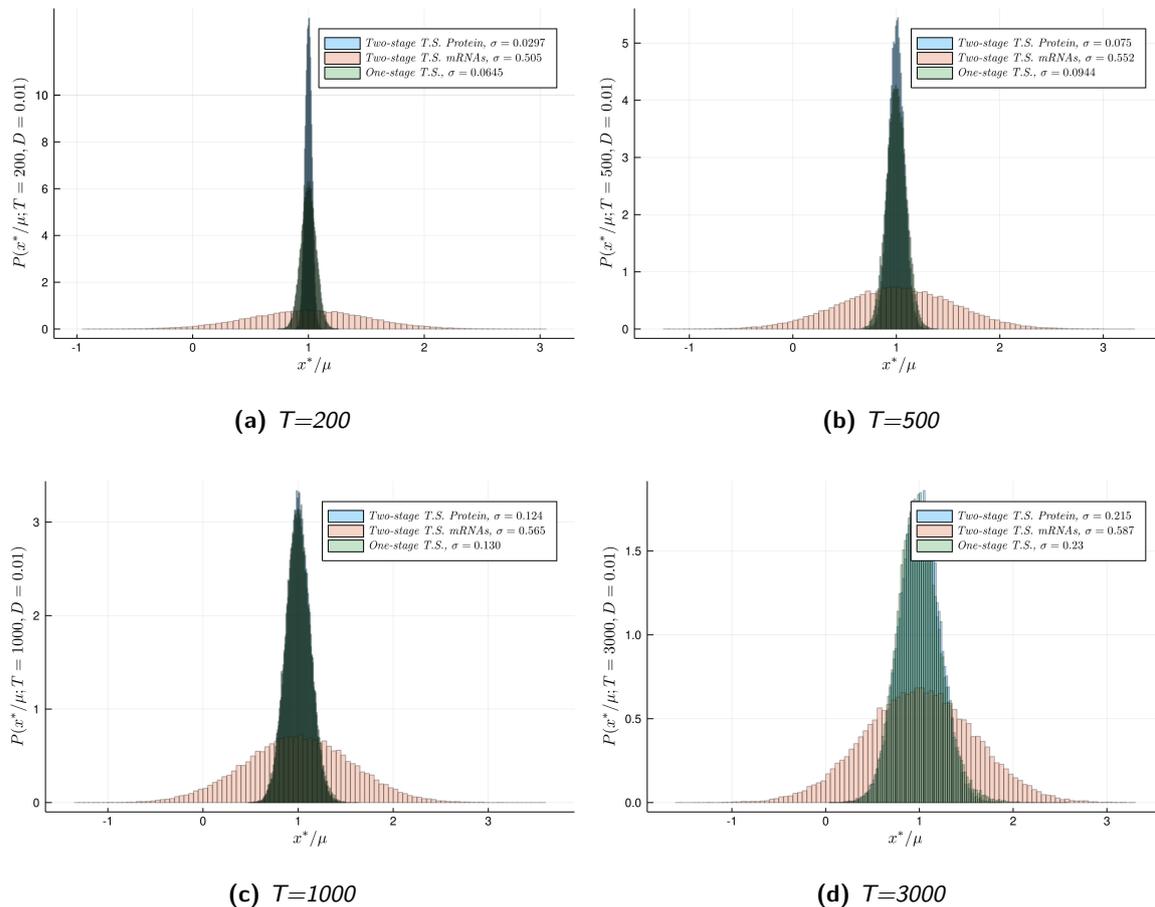
Figure 6.30: *Time evolution of the phenotype distributions of the 1-dimensional random variable x^*/μ referred to the protein concentration at two-stage T.S. (in cyan), for the mRNAs concentration at two-stage T.S. (in light pink) and the lumped level of both (in light green). The only difference to the previous analysis shown in Fig. 6.29 is the equality of $A = \alpha = 0.05$ and the corresponding time scales of the mRNA and protein level.*

Thus, the transcription level is purely dominated by the stochastic part of the dynamics and the consequent intrinsic dynamics variability generated from this noisy internal level is transferred to the phenotype level of regulation. All these results are in line with the biological evidence that the transcription level is dominated by molecular noise and the system is diffusive according to its gene dynamics and explores its isotropic local micro-state configurations [117]. This navigation is buffered by the translation level which modulates the mainly noisy-driven behavior of the transcription level. By tuning the noise, the system can perform an anisotropic exploration on the available space and make an biased tendency towards a specific

lineage in cell differentiation. Biologically speaking, this asymmetric noise might be placed by external input in the idea to represent the differentiation signals and drives the system towards one of its predefined attractors.

6.10.1 Population coupling effect on two-stage T.S. ensemble dynamics

Robustness and adaptability are the key processes to ensure biological survival [169]. Indeed, extra-cellular signaling can induce the cell to change its internal micro-state as a reaction to environmental perturbations. In other words, the cellular mechanisms, operating in a functional cell, make adaptive changes in the internal configuration as an active response to a multitude of factors. At the same time, the cell communicates to the outside by diverse mechanisms including the opening of ion channels, releasing membrane proteins, coupling to other receptors and releasing cytokines, microRNAs and transcription factors. Therefore, multicellular feedback loops can play a central role in cellular regulations.

Cells compose an inter-communicating colony that exchange nutrients, molecules and send out signals. Therefore the single action of one cell as a open system has to be understood in term of an individual reaction to a circulating flux of information at the population level. In this perspective, it is misleading, even in principle, to model biological systems by ignoring an information-based approach as a complementary analysis to address biological complexity [170, 171]. In particular, these intercellular mechanisms are further challenging the Waddington landscape as a quasi-equilibrium approach with a pre-defined potential which is actually originating from the dynamic multicellular self-organizing process. To incorporate this feedback loop control at the population level, I computationally studied a coupled ensemble of stochastic entities reflecting the two-stage T.S. model leading to a set of SDEs of the form

$$
\begin{cases}
\dot{X} = \alpha x - \gamma X + X_{eff} \\[2mm]
\dot{Y} = \alpha y - \gamma Y + Y_{eff} \\[2mm]
\dot{x} = A \frac{X+X_{eff}}{K_X+X+X_{eff}} \frac{K_X}{K_X+Y+Y_{eff}} - K^d x + \eta_X(t) \\[2mm]
\dot{y} = A \frac{Y+Y_{eff}}{K_Y+Y+Y_{eff}} \frac{K_Y}{K_Y+X+X_{eff}} - K^d y + \eta_Y(t) \,,
\end{cases}
\tag{6.32}
$$

with $\eta_X(t)$ and $\eta_Y(t)$ being random forces with the same statistical properties in analogy to Eq. 6.17 and the additional population meanfield coupling
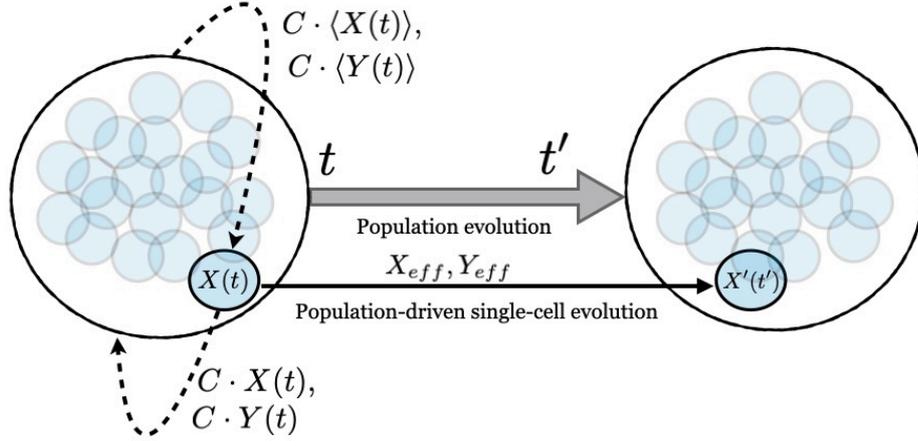
Figure 6.31: *Temporal evolution of the cell state $X(t)$ in a coupled dynamics with the entire population. The individual concentration is sent out at the population level and an average concentration signaling is received by each cell as an external global driving force.*

$$\begin{cases} X_{eff} = C(\langle X\rangle - X) \\ Y_{eff} = C(\langle Y\rangle - Y)\,, \end{cases} \tag{6.33}$$

where C denotes the coupling strength of the individual concentrations with the population average over time by a signaling loop (see Fig.6.31). When C is bigger than 1, the mean value sensed by each cell is amplified at the same way as the single concentration is communicated at the population level. This way of rearranging statistical information in a loop mechanism across scales, maintains the total mass invariant since $\langle X_{eff}(t)\rangle = 0$ and $\langle Y_{eff}(t)\rangle = 0$ at any time over simulation.

In order to study how the phenotype distributions would change under a population coupling, a stochastic ensemble was simulated with the parameter setting $\alpha = 0.05$, $\gamma = 0.0005$, $A = 0.5$, $K^d = 0.005$ and $D = 0.01$. As a first approach, I studied the symmetric case $K_X = K_Y = 2500$ with initial conditions $X_0 = 2650, Y_0 = 2450$, $x_0 = 26.5$ and $y_0 = 24.5$. The ensemble size was set to 4,000 and simulation time corresponded to $T=8000$. This computational study has the aim to study the evolution of the phenotype distributions until the system reaches a steady state. The result in Fig. 6.32 shows that in the uncoupled case ($C=0$) the standard deviations increase over time when the system is approaching a steady state configuration. Indeed, in Fig. 6.32a the phenotype distributions are overlapped and indistinguishable due to the protein concentration variability increasing over time. The dashed line indicates the steady protein concentration in respect to the initial condition. The dotted line indicates the mean of protein

concentration at time $T = 500$ for the phenotype distribution. In Fig. 6.32b the computational simulation result for a coupled population with $C=0.05$ is shown. Interestingly, the coupling effect at the population level is to stabilize the standard deviation over time. The phenotype distributions are narrowed around the mean, which changes over time by following the deterministic dynamics. This behavior becomes even more prominent by increasing C as shown in Fig. 6.32c where the standard deviation is stabilizing around 0.7.
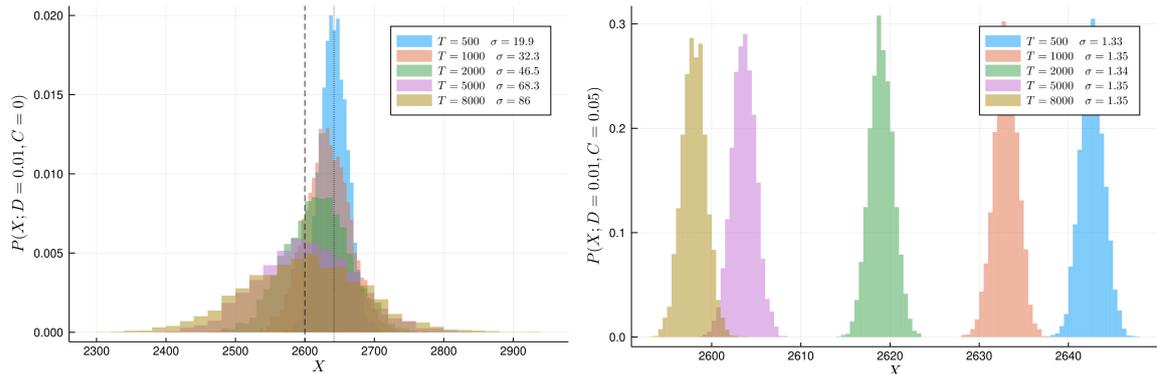
(a) *Dynamics for an uncoupled population.*

(b) *Dynamics for a coupled population with $C = 0.05$.*
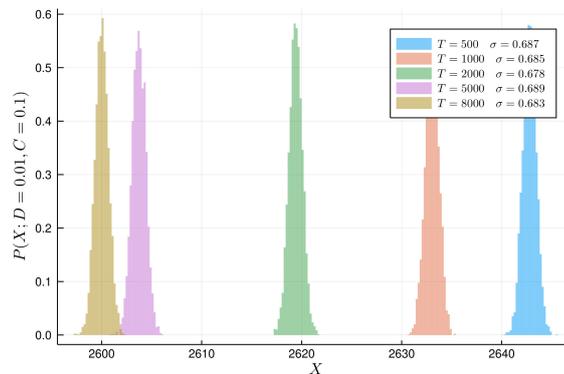
(c) *Dynamics for a coupled population with $C = 0.1$.*

Figure 6.32: *Dynamics of phenotype distributions, referred to X protein level, towards a steady state for a simulated ensemble of uncoupled (6.32a) and coupled (6.32b, 6.32c) cells. The particles starts from the concentrations $X_0 = 2650, Y_0 = 2450, x_0 = 26.5$ and $y_0 = 24.5$ and evolve under the SDEs in the symmetric two-stage S.T. with parameter setting: $\alpha = 0.05, \gamma = 0.0005, A = 0.5, K^d = 0.005$ and $D = 0.01$.*

In order to investigate the standard deviation behaviour in more detail, I studied its trends over time and across different C. The result is shown in Fig. 6.33 and corroborates the previous evidence of the decrease in variability of the phenotype distributions when a population coupling is introduced. The standard deviation
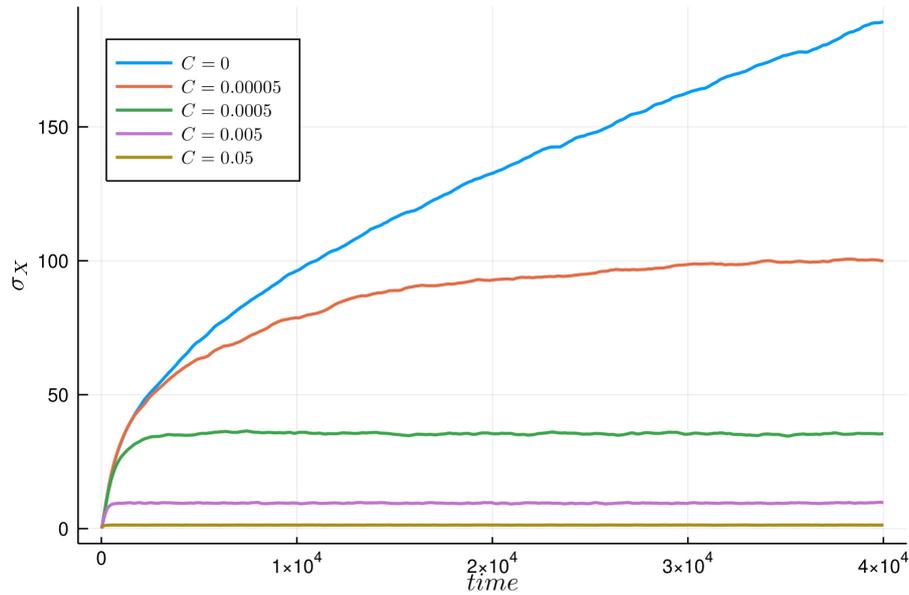
Figure 6.33: *Standard deviation trends over time in dependence on population coupling strength C.*

reaches a constant level after a period of time simulation depending on the coupling parameter C. Even with values rather small, such as 10^{-5}, the parameter C amplifies the additional information from the entire population to the single cell concentration over time. Therefore, the results in Figs. 6.32 and 6.33 indicates a self-adapted variability and directionality at the population level of the phenotype distribution in response to the meanfield interactions.

As a follow-up analysis, the uncoupled and coupled stochastic system dynamics on the stability manifold was studied. As the system reaches a steady state around $T = 8000$, I decided to evaluate the mean signal over a simulation until a final time $T = 40,000$. The result is shown in Fig. 6.34. When the system is approaching a steady state, the X protein concentration averaged over samples follows the deterministic trend independently on C. After an approximated time of $T = 8000$ when the system can be considered to be localized on the stability manifold, the mean signal of the X protein concentration starts to fluctuate in an apparently random matter in dependence on C.

106
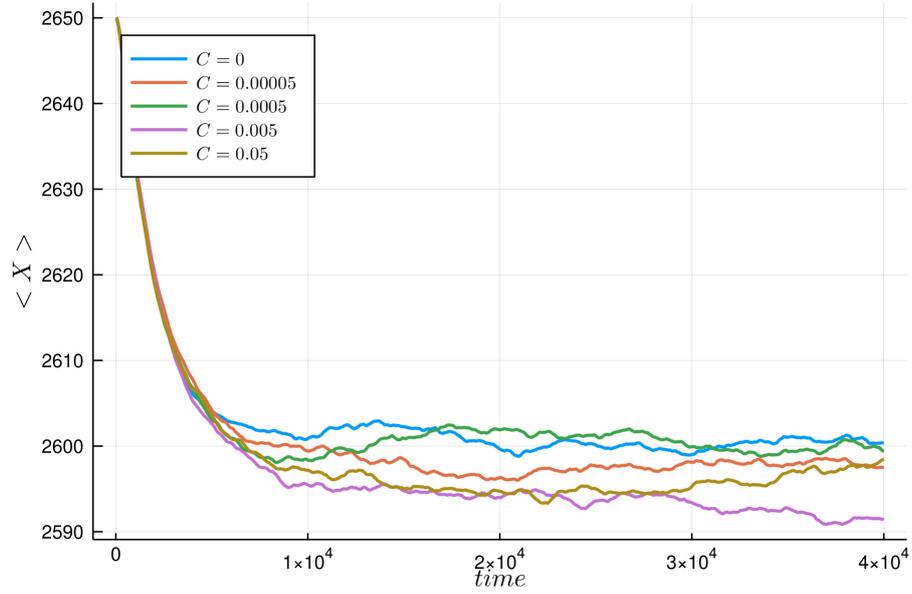
Figure 6.34: *Time evolution of the mean of the protein X concentration in dependence on coupling strengthe C.*

I finally characterized these fluctuations by averaging the standard deviation over time in comparison with averaging over the statistical ensemble. The aim is to study the effect of coupling on the ergodicity of the system [172] when it can freely move over a restricted phase space region associated to the stability manifold. Ergodicity refers to the idea that a point of a moving system, either driven by a deterministic or a stochastic process, will eventually visit all parts of the space that the system moves in, in an uniform and random sense. This implies, that the average behavior of the system can be deduced from the trajectory of a typical point. Referring to this specific analysis, the system holds the ergodic hypothesis if all the stochastic trajectories over the stationary manifold remain uniformly distributed at all the times. In order to verify this assessment, the standard deviation of the mean is chosen as a parameter of the system to evaluate the combined effect of statistical and temporal fluctuations. In the ergodic condition, the exchange between these two source of variability has to lead at the same outcome. I considered a long time window between $T = 8000$ and $T = 40000$ to study the statistical properties of the system that I quantified by (i) $\left\langle \sigma\left(X_M^{(t)}\right)\right\rangle_T$ and (ii) $\sigma\left(\left\langle X^{(m)}\right\rangle_T\right)_M$. The first quantity was evaluated by computing the standard deviation of the protein concentration over the statistical ensemble at any time and subsequently averaging the temporal signal over the time window. The second one was evaluated by performing the temporal mean of the protein concentration for any trajectory and consequently by computing the standard deviation of these values over the statistical ensemble. The result is on the Table 6.1.

To study the decreasing trend between the two evaluations across C, I introduced the quantity δ as

Coupling parameters	$\left\langle \sigma\!\left(X_M^{(t)}\right)\right\rangle_T$	$\sigma\!\left(\left\langle X^{(m)}\right\rangle_T\right)_M$
0	161	169
0.00005	86.1	98
0.0005	17	35.4
0.0007	13	30.3
0.005	2.20	9.54
0.05	0.231	1.34
0.1	0.11	0.688
0.3	0.038	0.23
0.5	0.023	0.140

Table 6.1: *Standard deviation of the mean related to the X protein concentration evaluated by averaging on statistical and temporal fluctuations across several population couplings.*

the ratio between $\sigma\!\left(\left\langle X^{(m)}\right\rangle_T\right)_M$ and $\left\langle \sigma\!\left(X_M^{(t)}\right)\right\rangle_T$. Finally, I plotted δ values over C.
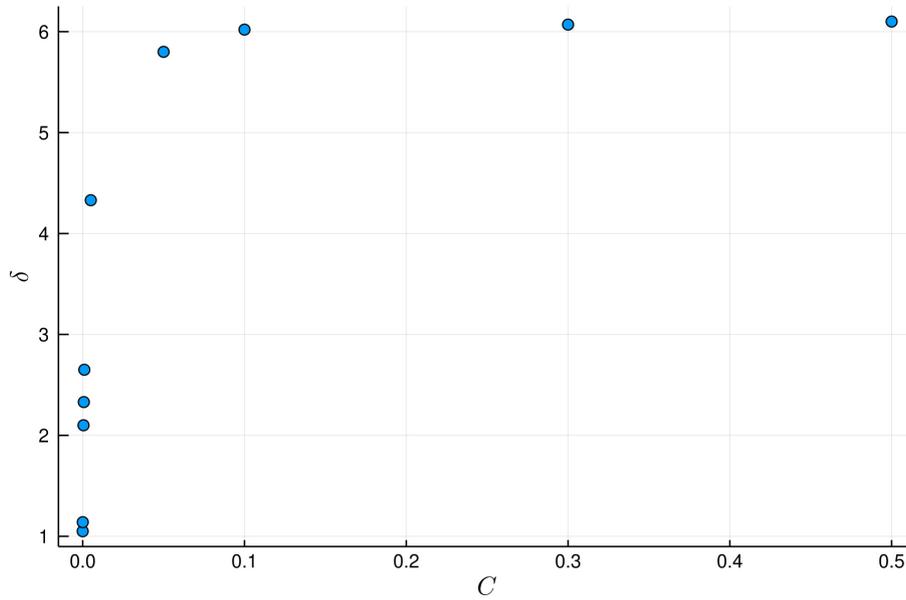
Figure 6.35: *Ratio between $\sigma\!\left(\left\langle X^{(m)}\right\rangle_T\right)_M$ and $\left\langle \sigma\!\left(X_M^{(t)}\right)\right\rangle_T$ over C.*

The outcome is reported in the Fig. 6.35 and indicates that δ becomes constant at a saturation level. At $C = 0$, $\sigma\!\left(\left\langle X^{(m)}\right\rangle_T\right)_M = \left\langle \sigma\!\left(X_M^{(t)}\right)\right\rangle_T$ is hold under statistical fluctuations. By increasing the population coupling C, the deviation from equality increases until a constant level is reached. Therefore the system deviates from the ergodic hypothesis upon modifying the statistical ensemble dynamics by adding a mean-interaction coupling at the protein level.

In his work [173], Boltzmann set up a more formal definition of the statistical ensembles as collections

of "monodes", i.e. probability distributions on phase space invariant under time evolution in a thermodynamic analogy. The main examples that are described are: the "holode", the two parameters collection (temperature and volume) of canonical distributions, and the "ergode", the two parameters (energy and volume) collection of microcanonical distributions. The ensembles are connected via equivalence properties and the ergodic hypothesis implies that the microcanonical distributions describe the statistics controlling the physics of a system modeled by given Hamiltonian microscopic equations. The hypothesis establishes the connection between physics and the thermodynamic analogies. Indeed "the ergodic hypothesis" in the sense of Boltzmann is that for a confined Hamiltonian system, "a phase space point evolves in time and eventually visit other points with the same energy" [174, 175]. In this prespectives, my final results might indicate that the system moves out-of-equilibrium driven by a meanfield coupling at the population level. The population coupling dynamics proposed above is a first naive attempt to model a stochastic dynamics driven by the population's signal. Therefore, further investigations with other coupling interactions are still necessary.

6.10.2 Conclusion on the symmetric two-stage toggle switch model

The extension of the model to the two regulatory levels of transcription and translation allowed for the analysis of the multi-level relation at the interface between genotype and phenotype. In particular, our analysis confirmed the essential role of the translation level to modulate the internal noise on the gene expression level. The buffer effect emerges at the slower time scale of the protein concentration compared to the mRNA production level. Overall, the dynamics is less dominated by noise leading to narrower phenotype distributions as shown in Fig. 6.29. This finding links to the general perspective that this internal mechanism of gene variability might provide cells with a self-regulation control mechanism upon stimuli and directionality during the cellular differentiation processes.

The coupling at the population level reduced even more the phenotype variability confirming that a positive feedback loop might produce an additional driving force towards one of the final stable states. Since the multi-particle meanfield interactions at the faster time scale affects the single-cell dynamics according to its microstate configuration, a predefined unique dynamics for all cells as supposed in the Waddington landascape framework is neglecting this self-organizational aspect of cell differentiation on the level of the population. Finally, I found that the statistical coupled ensemble does not respect the Boltzmann's ergodicity hypothesis [176] on the stability manifold and might indicate that the system is not equilibrating on the stationary manifold. This result indicates that the coupled cell population evolution, which describes a continuous cell-cell interactions over a stochastic meanfield-coupling dynamics, might induce a system in an out-of-equilibrium condition [177, 178, 179, 175].

6.11 Implications of a new perspective into modelling cell-fate decision

This chapter was dedicated to a bottom-up approach on cell fate dynamics. Based on the experimental results presented in Chapters 4 and 5 which indicated the importance of cellular diversity in cell inductions, I focused here on a mathematical framework and computational analysis to explore potential underlying mechanisms in toy models of differentiation. In Sections 6.6, 6.7, I introduced and studied for this purpose a model based on a one-stage genetic toggle switch scheme in a symmetric parameter setting. This analysis was performed from a deterministic and computational stochastic point of view to investigate the global and local properties of the systems dynamics. The model describes the stage of cellular differentiation characterized by a destabilization of the progenitor state that might correspond to a promiscuous metastable state poised between two possible alternative choices in the imminent cell-fate decision. The stability analysis combined with a stochastic investigation of the system allowed to reveal how the synergistic combination of deterministic dynamics and noise effects established peculiar heterogeneity for the molecular concentration probability distribution. Specifically, the noise accomplished that by allowing the system to reach neighboring steady states accordingly to the initial condition and its strength. Since the simulation analysis revealed that, accordingly to initial condition, the stationary probability distribution is localized on specific part of the stationary trajectory, the parameter K plays an interesting role as exploratory parameter on this restricted manifold. This property indicates the way of the system to loose partially its deterministic character and to gain a moderate flexibility in exploring other local configurations. This mechanistic behavior might reproduce the cellular tendency to assume diverse gene regulatory configurations during different stages of differentiation. Each of them is connected and/or referred to one of the two alternative lineages. Hence, in the assumption that the system can potentially explore other alternative steady states, in analogy to cells which investigate the possible internal microstates. Finally, the analysis shows that the noise allows the system to be aware of its overall dynamic properties even if a parameter choice was fixed.

External signals could break down this internal symmetry and direct the system towards a preferential stable state. This imbalance is implemented by structurally modifying the gene regulatory network and consequently cells are abruptly constrained to modify their internal configuration. Probabilistically, the external signal strengthens offers the possibility to visit those gene network configurations that are mainly affiliated to the differentiation phenotype profile. Accordingly, the inherent presence of noise and the instantaneous microstate configuration drive the cell-fate commitment. This aspect was investigated by studying the asymmetric one-stage toggle switch model (Section 6.9) again in the framework of distribution biology [27]. The noise is involved in determining the variation of the probability distribution evolution

over time. The results showed how the cellular heterogeneity might be emerged from the wide range of possibilities for each cells to reach the steady state. Indeed, the population dynamics indicates that, according to the initial molecular configurations, a characteristic variability over time is created by the coupling between noise and the asymmetric dynamics. From this perspective, a cell state distribution emerges from this mechanism of *induction* and *selection* at the population level.

A cell is also sensitive to external signals at larger scales of interactions as the macroscopic global behavior demonstrates e.g. in the local organization for tissue formation [180]. These long-range connections have to be considered in the modeling for cellular differentiation to respect multicellular signal propagation. An explicit analysis was performed in Section 6.10.1 demonstrating how an emerging behavior is originating by informing each cell about the instantaneous average from the entire population.

All these concepts have to be combined to improve the modeling description of cellular processes. The here presented modeling approaches, which are based on the symmetric and asymmetric toggle switch, can be directly applied to explore the mechanisms underlying the presence of "rebellious" cells during a cell fate induction process [39] as explained in Section 3.8. Note that the mechanistic model is not yet meant to provide a tool for direct applications in (sc-RNAseq) data integration but aims to open a new perspective in modelling cell fate decisions in a distribution biology framework.

Chapter 7

Conclusion and Outlook

This thesis is the result of my PhD study investigating cellular heterogeneity by a combined top-down analysis and bottom-up approach in the spirit of systems biology introduced in Fig. 3.1. This multidisciplinary study allowed me to acquire a broad knowledge of the available cutting-edge technologies and integrative frameworks tailored to address molecular and cellular variability. In the first part of the thesis, I applied an extended state-of-the-art data analysis tools [140] for single cell RNA-seq data to support the investigation of quantitative experiments dissecting and characterizing cellular heterogeneity and variable responses of the cells under different conditions and time courses. In this context, I focused on brain cell diversity and its relation to biomedical relevant systems. In the first study presented in Chapter 4 I investigated, in a collaborative project, the immune response of microglia as the immune cells of the brain to an activation stimulus. In the other interdisciplinary study of Chapter 5, I characterized the brain development in zebrafish during epileptic seizure establishment. These approaches are further discussed below in Section 7.1.

In the second part of the thesis (Chapter 6), I complemented the brain specific top-down approaches of the first part with a complementary and more generic bottom-up approach to investigate potential underlying mechanisms of cellular heterogeneity and cell fate. Thereby my gathered experience with the current challenges in biology and biomedicine, led me to the conclusion that coarse-grained models are essential to provide a complementary bottom-up perspective to understand cellular heterogeneity as an emergent property of underlying mechanisms which combine intrinsic gene expression variability and cell-cell interactions. For this purpose, I started from established theoretical approaches to describe cell fate [100] (in Sections 6.5,6.6), extended the framework and focused on a detailed analysis on the interplay between non-linear regulation and effects of the inherent molecular noise (Sections 6.7, **??**, 6.8 and 6.9). The impact and relation of this work to the first part of the thesis is discussed in more detail in Section 7.2.

7.1 Brain characterization at single cell resolution

The brain is the most complex organ in the body which involves billions of neurons for sending and receiving chemical and electrical signals throughout the body. Due to its high level of organization, many open questions still remain elusive and tricky to address. A major scientific interest in this context are clinical studies aiming at untangling brain disorders including neurodegeneration and epilepsy. The advancements in experimental approaches with single cell resolution like the single-cell RNA-sequencing technology (Section 3.3) have provided a complementary molecular-targeting approach to investigate biological processes. In particular in the study presented in Chapter 5, a detailed cell-population dissection of zebrafish brain development was performed by single-cell data analysis using specific gene markers. This detailed investigation confirmed the effect single mutations can have on differentiation processes and gave more insights into the current knowledge of epileptogenesis. In particular, a potential onset mechanism was hypothesised in accordance with the increase of reactive astrocytes [181] as a consequence for seizures and by structural modification of dendrites [182] observed in the fishline carrying the scn1a mutation.

In the more cell-type specific approach of Chapter 4, we investigated heterogeneity of cellular responses of microglia as the immune cells of the brain upon stimulation. This study is directly related to cell fate induction where external signals trigger a cellular adaptation in terms of physiological functions. In the case of microglia, this adaptation and potential hyper-activation is currently investigated in the context of different brain disorders where particularly for neurodegeneration such as in Parkinson's disease and Alzheimer's disease increasing evidence suggests that microglia can play a key role for neuronal death. From this perspective, our results on the induction of disease associated microglia (DAM) expression profiles and their heterogeneous appearance after stimulation may open new insights in underlying disease mechanisms. In particular, the detection of a delayed subpopulation provided molecular signatures for identifying possible defects in the microglia activation state. The elucidation of its heterogeneity might be also essential when considering its direct connections with neuronal cells [183] and its broad functions in the brain [184] including its role in neurodegeneration.

7.2 Mechanistic approaches to cell fate

Cellular differentiation and its intrinsic mechanism based on cell fate decision-making have been deeply studied from several perspectives [12, 185, 186, 187]). The most complete approaches are those that combine qualitative and quantitative investigations and attempt to integrate them across different time and spatial scales [11, 33]. The need for bridging these levels originate from the phenomenological evidence

of the complex multi-layers structure of multi-cellular organisms. Specifically, in the study of cellular differentiation, the cascade of events which regulates progressively cell specifications is modulated by both signal transduction pathways and cell-cell interactions. Zooming out from the detailed molecular behavior at small scales to a more extensive vision of multi-cellular dynamics, the Waddington epigenetic landscape laid the foundation for an abstract framework to globally capture cellular differentiation mechanisms [102]. Its way to look at the process as a dynamics evolving over subsequent quasi-static potentials, that are underpinned by gene regulatory network, has triggered fresh and innovative ideas.

Like any kind of models, also the Waddington potential landscape in its modern representation, has its own limitations and approximations. Indeed, this formulation is practically based on the hypothesis that a unique underlying deterministic dynamics governs the process from the beginning. It assumes that the system is somehow aware of preferential directions to be taken along the process and its abrupt changes are justified by a time-dependent parameter of the dynamics which establishes critical transitions that eventually put the system in another stability regime. I consider this as a strong basic assumption in the model. From my perspective, the internal changes of gene interactions might occur in many unpredictable ways depending on a plethora of factors. This consideration led me to the conclusion that a more flexible modelling approach for cell fate induction is necessary. Moreover, due to the persistent out-of-equilibrium condition of cells, they are continuously stimulated and in exchange with their environment to adapt themselves to diverse stimuli. Hence, the quasi-potential approximation might be conceptually too far from representing reality.

Here, I proposed a more general viewpoint to model cellular differentiation by introducing a symmetric one stage toggle-switch model in Section 6.6 and its asymmetric version in Section 6.9 to mechanistically link together the two distinct stages of the process which conduct the cell system towards a final cell state configuration (Fig. 2.1). I assume that cell commitment is deemed as an individual choice performed by induction-selection mechanism between the internal gene expression and external signals in terms of a cellular nano-evolution [27]. I extended this concept inside the biology distribution framework (Section 3.8) where gene expression variability is the intrinsic stochastic source to create cellular heterogeneity as a macroscopic manifestation of a stochastic multicellular dynamics as described in Sections 6.8 and 6.9). Furthermore, the Waddington landscape's perspective does not consider the cell-cell interactions, which play a central role during cell commitment and consequently in the final cell state at the population level as an emergent phenomenon. In Section 6.10.1, I added a population coupling in a two stage toggle-switch model and by qualitatively studying the stochastic dynamics, where a collective effect could be identified. That demonstrates that a feedback control mechanism is able to create an additional driving force at the population level in the context of an independent stochastic dynamics, which pushed the cells towards

114

their final phenotype distribution. This finding convinces me about the stringent requirement to consider a larger scale phenomena as a global manifestation of many-interacting interactions at smaller scales. For model validation and fine-tuning, it is urgent to collect and combine data from different levels by diverse complementary approaches. This information flux and integrative knowledge is the way to guide theoretical approaches to fill experimental gaps and potentially provide a more universal understanding of nature.

7.3   Interdisciplinary approaches to address the urgent challenges

Although inferring specific information from data and comparing qualitative and quantitative measurements is the correct approach to acquire reliable clues, a systematic way to integrate multi-omics data is still lacking [7]. One reason for this gap is the huge amount of information to be cross-validated. With the aim to provide a complementary approach to address this challenge by System Biology, theoretical bottom-up investigations have to be posed to facilitate the bridging over different scales to describe how biological information is propagated from the genotype to phenotype level to sustain the organism as a whole. Therefore, a better understanding for the general concept of co-existence of robustness and adaptability in regulated biological system is needed [188]. I attempted to put all these main concepts into a simple but generic model for studying cellular differentiation with the idea to understand the basic mechanisms as emergent behavior from internal variability combined with direct instructions [20].

Based on its basic and general features, the model has to be adapted for the analysis of specific binary cell fates by validating its corresponding predictions and identifying its limitations. During this process and tuned modifications, the model might provide insights into analysing and interpreting specific sc-RNAseq data based on the identified proposed underlying mechanisms for the emergence of cellular heterogeneity as a manifestation of dynamic phenotype distributions.

Overall, the experience during my PhD period has strengthened my believe that interdisciplinarity as a collaboration of different expertise is the only way to achieve a comprehensive and deep understanding of any natural and social phenomenon and to address the urgent challenges of our society. In this regard, I have actively been part of the Research Luxembourg COVID-19 Task force, which was founded to support the Luxembourg government in managing the pandemic crisis. This interdisciplinary group of medical doctors, virologists, epidemiologists, mathematicians, physicists and other researchers has demonstrated how solid interdisciplinary scientific approaches can initiated fact-based discussions and decisions to eventually support the society.

References

[1] R. Schrodinger, E. Schrödinger and E. S. Dinger. *What is life?: With mind and matter and auto-biographical sketches.* Cambridge university press, 1992.

[2] S. Pilosof et al. 'The multilayer nature of ecological networks'. In: *Nature Ecology & Evolution* 1.4 (2017), pp. 1–9.

[3] L. G. Alves et al. 'The nested structural organization of the worldwide trade multi-layer network'. In: *Scientific reports* 9.1 (2019), pp. 1–14.

[4] R. S. Burt. 'Models of network structure'. In: *Annual review of sociology* 6.1 (1980), pp. 79–141.

[5] Z. Hammoud and F. Kramer. 'Multilayer networks: aspects, implementations, and application in biomedicine'. In: *Big Data Analytics* 5.1 (2020), pp. 1–18.

[6] L. V. Schaffer and T. Ideker. 'Mapping the multiscale structure of biological systems'. In: *Cell Systems* 12.6 (2021), pp. 622–635.

[7] M. P. Nagle et al. 'Bridging scales: From cell biology to physiology using in situ single-cell technologies'. In: *Cell Systems* 12.5 (2021), pp. 388–400.

[8] M. Rauzi et al. 'Embryo-scale tissue mechanics during Drosophila gastrulation movements'. In: *Nature communications* 6.1 (2015), pp. 1–12.

[9] S. Wennekamp et al. 'A self-organization framework for symmetry breaking in the mammalian embryo'. In: *Nature reviews Molecular cell biology* 14.7 (2013), pp. 452–459.

[10] K. W. Rogers and A. F. Schier. 'Morphogen gradients: from generation to interpretation'. In: *Annual review of cell and developmental biology* 27 (2011), pp. 377–407.

[11] N. Folguera-Blasco et al. 'A multiscale model of epigenetic heterogeneity-driven cell fate decision-making'. In: *PLoS computational biology* 15.4 (2019), e1006592.

[12] T. M. Norman et al. 'Memory and modularity in cell-fate decision making'. In: *Nature* 503.7477 (2013), pp. 481–486.

[13] E. B. Dewey, D. T. Taylor and C. A. Johnston. 'Cell fate decision making through oriented cell division'. In: *Journal of developmental biology* 3.4 (2015), pp. 129–157.

[14] X.-P. Zhang et al. 'Cell fate decision mediated by p53 pulses'. In: *Proceedings of the National Academy of Sciences* 106.30 (2009), pp. 12245–12250.

[15] B. D. MacArthur, A. Ma'ayan and I. R. Lemischka. 'Systems biology of stem cell fate and cellular reprogramming'. In: *Nature reviews Molecular cell biology* 10.10 (2009), pp. 672–681.

[16] P. W. Anderson. 'More is different'. In: *Science* 177.4047 (1972), pp. 393–396.

[17] N. Guisoni et al. 'Diversity of fate outcomes in cell pairs under lateral inhibition'. In: *Development* 144.7 (2017), pp. 1177–1186.

[18] B. Egger, K. S. Gold and A. H. Brand. 'Notch regulates the switch from symmetric to asymmetric neural stem cell division in the Drosophila optic lobe'. In: *Development* 137.18 (2010), pp. 2981–2987.

[19] F. M. Watt. 'Engineered microenvironments to direct epidermal stem cell behavior at single-cell resolution'. In: *Developmental Cell* 38.6 (2016), pp. 601–609.

[20] A. Xavier da Silveira dos Santos and P. Liberali. 'From single cells to tissue self-organization'. In: *The FEBS Journal* 286.8 (2019), pp. 1495–1513.

[21] Q. Chen et al. 'Tracing the origin of heterogeneity and symmetry breaking in the early mammalian embryo'. In: *Nature communications* 9.1 (2018), pp. 1–11.

[22] D. M. Gordon. 'The ecology of collective behavior'. In: *PLoS biology* 12.3 (2014), e1001805.

[23] J. Elgeti, R. G. Winkler and G. Gompper. 'Physics of microswimmerssingle particle motion and collective behavior: a review'. In: *Reports on progress in physics* 78.5 (2015), p. 056601.

[24] W. Wang et al. 'From one to many: Dynamic assembly and collective behavior of self-propelled colloidal motors'. In: *Accounts of chemical research* 48.7 (2015), pp. 1938–1946.

[25] D. Patra et al. 'Intelligent, self-powered, drug delivery systems'. In: *Nanoscale* 5.4 (2013), pp. 1273–1283.

[26] T. S. Deisboeck and I. D. Couzin. 'Collective behavior in cancer cell populations'. In: *Bioessays* 31.2 (2009), pp. 190–197.

[27] N. Komin and A. Skupin. 'How to address cellular heterogeneity by distribution biology'. In: *Current Opinion in Systems Biology* 3 (2017), pp. 154–160.

[28] C. Schuurmans and F. Guillemot. 'Molecular mechanisms underlying cell fate specification in the developing telencephalon'. In: *Current opinion in neurobiology* 12.1 (2002), pp. 26–34.

[29] E. H. Davidson. 'How embryos work: a comparative view of diverse modes of cell fate specification'. In: *Development* 108.3 (1990), pp. 365–389.

[30] L. Vallier et al. 'Early cell fate decisions of human embryonic stem cells and mouse epiblast stem cells are controlled by the same signalling pathways'. In: *PloS one* 4.6 (2009), e6082.

[31] A. Guillemin and M. P. Stumpf. 'Noise and the molecular processes underlying cell fate decision-making'. In: *Physical biology* 18.1 (2020), p. 011002.

[32] T. S. Gardner, C. R. Cantor and J. J. Collins. 'Construction of a genetic toggle switch in Escherichia coli'. In: *Nature* 403.6767 (2000), pp. 339–342.

[33] M. Strasser, F. J. Theis and C. Marr. 'Stability and multiattractor dynamics of a toggle switch based on a two-stage model of stochastic gene expression'. In: *Biophysical journal* 102.1 (2012), pp. 19–29.

[34] D. T. Gillespie. 'Exact stochastic simulation of coupled chemical reactions'. In: *The journal of physical chemistry* 81.25 (1977), pp. 2340–2361.

[35] S. L. Spencer et al. 'Non-genetic origins of cell-to-cell variability in TRAIL-induced apoptosis'. In: *Nature* 459.7245 (2009), pp. 428–432.

[36] L. T. MacNeil and A. J. Walhout. 'Gene regulatory networks and the role of robustness and stochasticity in the control of gene expression'. In: *Genome research* 21.5 (2011), pp. 645–657.

[37] S. S. Shen-Orr et al. 'Network motifs in the transcriptional regulation network of Escherichia coli'. In: *Nature genetics* 31.1 (2002), pp. 64–68.

[38] U. Alon. 'Network motifs: theory and experimental approaches'. In: *Nature Reviews Genetics* 8.6 (2007), pp. 450–461.

[39] M. Mojtahedi et al. 'Cell fate decision as high-dimensional critical state transition'. In: *PLoS biology* 14.12 (2016), e2000640.

[40] E. Bianconi et al. 'An estimation of the number of cells in the human body'. In: *Annals of human biology* 40.6 (2013), pp. 463–471.

[41] H. Kitano. 'Systems biology: a brief overview'. In: *science* 295.5560 (2002), pp. 1662–1664.

[42] T. Höfer, L. Venance and C. Giaume. 'Control and plasticity of intercellular calcium waves in astrocytes: a modeling approach'. In: *Journal of Neuroscience* 22.12 (2002), pp. 4850–4859.

[43] S. Tyanova et al. 'The Perseus computational platform for comprehensive analysis of (prote) omics data'. In: *Nature methods* 13.9 (2016), pp. 731–740.

[44] S. B. Kotsiantis, I. D. Zaharakis and P. E. Pintelas. 'Machine learning: a review of classification and combining techniques'. In: *Artificial Intelligence Review* 26.3 (2006), pp. 159–190.

[45] D. G. Spiller et al. 'Measurement of single-cell dynamics'. In: *Nature* 465.7299 (2010), pp. 736–745.

[46] D. Türei et al. 'Integrated intra-and intercellular signaling knowledge for multicellular omics analysis'. In: *Molecular systems biology* 17.3 (2021), e9923.

[47] T. Stuart and R. Satija. 'Integrative single-cell analysis'. In: *Nature Reviews Genetics* 20.5 (2019), pp. 257–272.

[48] A. Mortazavi et al. 'Mapping and quantifying mammalian transcriptomes by RNA-Seq'. In: *Nature methods* 5.7 (2008), pp. 621–628.

[49] J. C. Marioni et al. 'RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays'. In: *Genome research* 18.9 (2008), pp. 1509–1517.

[50] Z. Wang, M. Gerstein and M. Snyder. 'RNA-Seq: a revolutionary tool for transcriptomics'. In: *Nature reviews genetics* 10.1 (2009), pp. 57–63.

[51] D. Grün, L. Kester and A. Van Oudenaarden. 'Validation of noise models for single-cell transcriptomics'. In: *Nature methods* 11.6 (2014), pp. 637–640.

[52] O. Stegle, S. A. Teichmann and J. C. Marioni. 'Computational and analytical challenges in single-cell transcriptomics'. In: *Nature Reviews Genetics* 16.3 (2015), pp. 133–145.

[53] E. Pierson and C. Yau. 'ZIFA: Dimensionality reduction for zero-inflated single-cell gene expression analysis'. In: *Genome biology* 16.1 (2015), pp. 1–10.

[54] J. K. Kim and J. C. Marioni. 'Inferring the kinetics of stochastic gene expression from single-cell RNA-sequencing data'. In: *Genome biology* 14.1 (2013), pp. 1–12.

[55] M. Kaern et al. 'Stochasticity in gene expression: from theories to phenotypes'. In: *Nature Reviews Genetics* 6.6 (2005), pp. 451–464.

[56] A. Raj and A. Van Oudenaarden. 'Nature, nurture, or chance: stochastic gene expression and its consequences'. In: *Cell* 135.2 (2008), pp. 216–226.

[57]   V. Marx. 'Method of the Year: spatially resolved transcriptomics'. In: *Nature Methods* 18.1 (2021), pp. 9–14.

[58]   K. Takahashi and S. Yamanaka. 'Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors'. In: *cell* 126.4 (2006), pp. 663–676.

[59]   G. F. Merrill. 'Cell synchronization'. In: *Methods in cell biology* 57 (1998), pp. 229–249.

[60]   A. Balsalobre, F. Damiola and U. Schibler. 'A serum shock induces circadian gene expression in mammalian tissue culture cells'. In: *Cell* 93.6 (1998), pp. 929–937.

[61]   R. Métivier et al. 'Estrogen receptor-α directs ordered, cyclical, and combinatorial recruitment of cofactors on a natural target promoter'. In: *Cell* 115.6 (2003), pp. 751–763.

[62]   L. Ashall et al. 'Pulsatile stimulation determines timing and specificity of NF-κB-dependent transcription'. In: *Science* 324.5924 (2009), pp. 242–246.

[63]   F. Buettner et al. 'Computational analysis of cell-to-cell heterogeneity in single-cell RNA-sequencing data reveals hidden subpopulations of cells'. In: *Nature biotechnology* 33.2 (2015), pp. 155–160.

[64]   J. Cao et al. 'Sci-fate characterizes the dynamics of gene expression in single cells'. In: *Nature biotechnology* 38.8 (2020), pp. 980–988.

[65]   A. H. Rizvi et al. 'Single-cell topological RNA-seq analysis reveals insights into cellular differentiation and development'. In: *Nature biotechnology* 35.6 (2017), pp. 551–560.

[66]   G. La Manno et al. 'RNA velocity of single cells'. In: *Nature* 560.7719 (2018), pp. 494–498.

[67]   M. Setty et al. 'Wishbone identifies bifurcating developmental trajectories from single-cell data'. In: *Nature biotechnology* 34.6 (2016), pp. 637–645.

[68]   G. Karlebach and R. Shamir. 'Modelling and analysis of gene regulatory networks'. In: *Nature reviews Molecular cell biology* 9.10 (2008), pp. 770–780.

[69]   X. Qiu et al. 'Towards inferring causal gene regulatory networks from single cell expression measurements'. In: *BioRxiv* (2018), p. 426981.

[70]   M. W. Fiers et al. 'Mapping gene regulatory networks from single-cell omics data'. In: *Briefings in functional genomics* 17.4 (2018), pp. 246–254.

[71]   A. A. Margolin et al. 'ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context'. In: *BMC bioinformatics*. Vol. 7. 1. Springer. 2006, pp. 1–15.

[72] S. Zhou et al. 'Single-cell RNA-seq dissects the intratumoral heterogeneity of triple-negative breast cancer based on gene regulatory networks'. In: *Molecular Therapy-Nucleic Acids* 23 (2021), pp. 682–690.

[73] J. Wang et al. 'Gene expression distribution deconvolution in single-cell RNA sequencing'. In: *Proceedings of the National Academy of Sciences* 115.28 (2018), E6437–E6446.

[74] J. Ahn et al. 'DeMix: deconvolution for mixed cancer transcriptomes using raw measured data'. In: *Bioinformatics* 29.15 (2013), pp. 1865–1871.

[75] H. Jiang et al. 'Single cell clustering based on cell-pair differentiability correlation and variance analysis'. In: *Bioinformatics* 34.21 (2018), pp. 3684–3694.

[76] T. M. Ignac et al. 'Discovering pair-wise genetic interactions: an information theory-based approach'. In: *PloS one* 9.3 (2014), e92310.

[77] B. Huang et al. 'Interrogating the topological robustness of gene regulatory circuits by randomization'. In: *PLoS computational biology* 13.3 (2017), e1005456.

[78] J. Casadiego et al. 'Model-free inference of direct network interactions from nonlinear collective dynamics'. In: *Nature communications* 8.1 (2017), pp. 1–10.

[79] H. Ma et al. 'Randomly distributed embedding making short-term high-dimensional data predictable'. In: *Proceedings of the National Academy of Sciences* 115.43 (2018), E9994–E10002.

[80] C. J. Gregory and A. C. Eaves. 'Human marrow cells capable of erythropoietic differentiation in vitro: definition of three erythroid colony responses'. In: (1977).

[81] J. S. Rush and P. D. Hodgkin. 'B cells activated via CD40 and IL-4 undergo a division burst but require continued stimulation to maintain division, survival and differentiation'. In: *European journal of immunology* 31.4 (2001), pp. 1150–1159.

[82] N. Planque. 'Nuclear trafficking of secreted factors and cell-surface receptors: new pathways to regulate cell proliferation and differentiation, and involvement in cancers'. In: *Cell Communication and Signaling* 4.1 (2006), pp. 1–18.

[83] A. Richard et al. 'Single-cell-based analysis highlights a surge in cell-to-cell molecular variability preceding irreversible commitment in a differentiation process'. In: *PLoS biology* 14.12 (2016), e1002585.

[84] L. Kester and A. van Oudenaarden. 'Single-cell transcriptomics meets lineage tracing'. In: *Cell stem cell* 23.2 (2018), pp. 166–179.

[85] V. Moignard et al. 'Characterization of transcriptional networks in blood stem and progenitor cells using high-throughput single-cell gene expression analysis'. In: *Nature cell biology* 15.4 (2013), pp. 363–372.

[86] M. B. Elowitz et al. 'Stochastic gene expression in a single cell'. In: *Science* 297.5584 (2002), pp. 1183–1186.

[87] D. M. Suter et al. 'Mammalian genes are transcribed with widely different bursting kinetics'. In: *science* 332.6028 (2011), pp. 472–474.

[88] A. Raj et al. 'Imaging individual mRNA molecules using multiple singly labeled probes'. In: *Nature methods* 5.10 (2008), pp. 877–879.

[89] I. Lestas, G. Vinnicombe and J. Paulsson. 'Fundamental limits on the suppression of molecular fluctuations'. In: *Nature* 467.7312 (2010), pp. 174–178.

[90] J. Viñuelas et al. 'Towards experimental manipulation of stochasticity in gene expression'. In: *Progress in biophysics and molecular biology* 110.1 (2012), pp. 44–53.

[91] D. Huh and J. Paulsson. 'Non-genetic heterogeneity from stochastic partitioning at cell division'. In: *Nature genetics* 43.2 (2011), pp. 95–100.

[92] T. Çaatay et al. 'Architecture-dependent noise discriminates functionally analogous differentiation circuits'. In: *Cell* 139.3 (2009), pp. 512–522.

[93] J. Viñuelas et al. 'Quantifying the contribution of chromatin dynamics to stochastic gene expression reveals long, locus-dependent periods between transcriptional bursts'. In: *BMC biology* 11.1 (2013), pp. 1–19.

[94] A. Raj et al. 'Stochastic mRNA synthesis in mammalian cells'. In: *PLoS biology* 4.10 (2006), e309.

[95] D. J. Wilkinson. *Stochastic modelling for systems biology*. Chapman and Hall/CRC, 2018.

[96] A. Tosenberger et al. 'A multiscale model of early cell lineage specification including cell division'. In: *NPJ systems biology and applications* 3.1 (2017), pp. 1–11.

[97] Z. Cang et al. 'A multiscale model via single-cell transcriptomics reveals robust patterning mechanisms during early mammalian embryo development'. In: *PLoS computational biology* 17.3 (2021), e1008571.

[98] J.-J. Kupiec. 'A Darwinian theory for the origin of cellular differentiation'. In: *Molecular and General Genetics MGG* 255.2 (1997), pp. 201–208.

[99]   D. V. Foster et al. 'A model of sequential branching in hierarchical cell fate determination'. In: *Journal of theoretical biology* 260.4 (2009), pp. 589–597.

[100] S. Huang et al. 'Cell fates as high-dimensional attractor states of a complex gene regulatory network'. In: *Physical review letters* 94.12 (2005), p. 128701.

[101] T. Kai and A. Spradling. 'Differentiating germ cells can revert into functional stem cells in Drosophila melanogaster ovaries'. In: *Nature* 428.6982 (2004), pp. 564–569.

[102] N. Moris, C. Pina and A. M. Arias. 'Transition states and cell fate decisions in epigenetic landscapes'. In: *Nature Reviews Genetics* 17.11 (2016), pp. 693–703.

[103] J. De Caluwé et al. 'Signalling-modulated gene regulatory networks in early mammalian development'. In: *Journal of theoretical biology* 463 (2019), pp. 56–66.

[104] L. Bast et al. 'Computational modeling of stem and progenitor cell kinetics identifies plausible hematopoietic lineage hierarchies'. In: *Iscience* 24.2 (2021), p. 102120.

[105] S. Huang. 'Systems biology of stem cells: three useful perspectives to help overcome the paradigm of linear pathways'. In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 366.1575 (2011), pp. 2247–2259.

[106] C. H. Waddington. 'Canalization of development and the inheritance of acquired characters'. In: *Nature* 150.3811 (1942), pp. 563–565.

[107] C. H. Waddington. 'Towards a theoretical biology'. In: *Nature* 218.5141 (1968), pp. 525–527.

[108] C. H. Waddington. *The strategy of the genes*. Routledge, 2014.

[109] S. Huang. 'The molecular and mathematical basis of Waddington's epigenetic landscape: A framework for post-Darwinian biology?' In: *Bioessays* 34.2 (2012), pp. 149–157.

[110] J. Wang et al. 'The potential landscape of genetic circuits imposes the arrow of time in stem cell differentiation'. In: *Biophysical journal* 99.1 (2010), pp. 29–39.

[111] J. E. Marsden and A. Tromba. *Vector calculus*. Macmillan, 2003.

[112] E. Penocchio, R. Rao and M. Esposito. 'Thermodynamic efficiency in dissipative chemistry'. In: *Nature communications* 10.1 (2019), pp. 1–5.

[113] R. Rao and M. Esposito. 'Nonequilibrium thermodynamics of chemical reaction networks: wisdom from stochastic thermodynamics'. In: *Physical Review X* 6.4 (2016), p. 041064.

[114] C. Schröter et al. 'FGF/MAPK signaling sets the switching threshold of a bistable circuit controlling cell fate decisions in embryonic stem cells'. In: *Development* 142.24 (2015), pp. 4205–4216.

[115] R. Ahrends et al. 'Controlling low rates of cell differentiation through noise and ultrahigh feedback'. In: *Science* 344.6190 (2014), pp. 1384–1389.

[116] S. Huang et al. 'Bifurcation dynamics in lineage-commitment in bipotent progenitor cells'. In: *Developmental biology* 305.2 (2007), pp. 695–713.

[117] J. Trott et al. 'Dissecting ensemble networks in ES cell populations reveals micro-heterogeneity underlying pluripotency'. In: *Molecular BioSystems* 8.3 (2012), pp. 744–752.

[118] C. Marr, J. X. Zhou and S. Huang. 'Single-cell gene expression profiling and cell state dynamics: collecting data, correlating data points and connecting the dots'. In: *Current opinion in biotechnology* 39 (2016), pp. 207–214.

[119] R. L. Coffman and S. L. Reiner. 'Instruction, selection, or tampering with the odds?' In: *Science* 284.5418 (1999), pp. 1283–1285.

[120] R. N. Germain. 'T-cell development and the CD4–CD8 lineage decision'. In: *Nature reviews immunology* 2.5 (2002), pp. 309–322.

[121] L. Robb. 'Cytokine receptors and hematopoietic differentiation'. In: *Oncogene* 26.47 (2007), pp. 6715–6723.

[122] T. Graf. 'Differentiation plasticity of hematopoietic cells'. In: *Blood, The Journal of the American Society of Hematology* 99.9 (2002), pp. 3089–3101.

[123] T. Enver and S. E. W. Jacobsen. 'Instructions writ in blood'. In: *Nature* 461.7261 (2009), pp. 183–184.

[124] T. Dobzhansky. 'Nothing in biology makes sense except in the light of evolution'. In: *The american biology teacher* 75.2 (2013), pp. 87–91.

[125] S. Huang. 'Cell lineage determination in state space: a systems view brings flexibility to dogmatic canonical rules'. In: *PLoS biology* 8.5 (2010), e1000380.

[126] S. Huang. 'Non-genetic heterogeneity of cells in development: more than just noise'. In: *Development* 136.23 (2009), pp. 3853–3862.

[127] G. Dong et al. 'Resilience of networks with community structure behaves as if under an external field'. In: *Proceedings of the National Academy of Sciences* 115.27 (2018), pp. 6911–6915.

[128] M. E. Gáspár and P. Csermely. 'Rigidity and flexibility of biological networks'. In: *Briefings in Functional Genomics* 11.6 (2012), pp. 443–456.

[129] J. Garcia-Ojalvo and A. M. Arias. 'Towards a statistical mechanics of cell fate decisions'. In: *Current opinion in genetics & development* 22.6 (2012), pp. 619–626.

[130] Z. Tan et al. 'Aneuploidy underlies a multicellular phenotypic switch'. In: *Proceedings of the National Academy of Sciences* 110.30 (2013), pp. 12367–12372.

[131] A. Skupin, H. Kettenmann and M. Falcke. 'Calcium signals driven by single channel noise'. In: *PLoS computational biology* 6.8 (2010), e1000870.

[132] C. Sousa et al. 'Single-cell transcriptomics reveals distinct inflammation-induced microglia signatures'. In: *EMBO reports* 19.11 (2018), e46171.

[133] A. London, M. Cohen and M. Schwartz. 'Microglia and monocyte-derived macrophages: functionally distinct populations that act in concert in CNS plasticity and repair'. In: *Frontiers in cellular neuroscience* 7 (2013), p. 34.

[134] H. Keren-Shaul et al. 'A unique microglia type associated with restricting development of Alzheimers disease'. In: *Cell* 169.7 (2017), pp. 1276–1290.

[135] W. Zhang et al. 'Aggregated α-synuclein activates microglia: a process leading to disease progression in Parkinson's disease'. In: *The FASEB Journal* 19.6 (2005), pp. 533–542.

[136] E. Z. Macosko et al. 'Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets'. In: *Cell* 161.5 (2015), pp. 1202–1214.

[137] M. D. Luecken and F. J. Theis. 'Current best practices in single-cell RNA-seq analysis: a tutorial'. In: *Molecular systems biology* 15.6 (2019), e8746.

[138] X. Qiu et al. 'Single-cell mRNA quantification and differential analysis with Census'. In: *Nature methods* 14.3 (2017), pp. 309–315.

[139] E. Tiraboschi et al. 'New insights into the early mechanisms of epileptogenesis in a zebrafish model of Dravet syndrome'. In: *Epilepsia* 61.3 (2020), pp. 549–560.

[140] C. Trapnell and D. Cacchiarelli. 'Monocle: Differential expression and time-series analysis for single-cell RNA-Seq and qPCR experiments'. In: *Bioconductor. Fmrp. Usp. Br* 2 (2014).

[141] B. G. Murray. 'Universal laws and predictive theory in ecology and evolution'. In: *Oikos* 89.2 (2000), pp. 403–408.

[142] Z.-S. She and Z.-X. Zhang. 'Universal hierarchical symmetry for turbulence and general multi-scale fluctuation systems'. In: *Acta Mechanica Sinica* 25.3 (2009), pp. 279–294.

[143] H. H. Chang et al. 'Multistable and multistep dynamics in neutrophil differentiation'. In: *BMC cell biology* 7.1 (2006), pp. 1–12.

[144] J.-B. Lugagne et al. 'Balancing a genetic toggle switch by real-time feedback control and periodic forcing'. In: *Nature communications* 8.1 (2017), pp. 1–8.

[145] H. Kobayashi et al. 'Programmable cells: interfacing natural and engineered gene networks'. In: *Proceedings of the National Academy of Sciences* 101.22 (2004), pp. 8414–8419.

[146] V. Mutzel et al. 'A symmetric toggle switch explains the onset of random X inactivation in different mammals'. In: *Nature structural & molecular biology* 26.5 (2019), pp. 350–360.

[147] M. Scheffer et al. 'Early-warning signals for critical transitions'. In: *Nature* 461.7260 (2009), pp. 53–59.

[148] T. Tian and K. Burrage. 'Stochastic models for regulatory networks of the genetic toggle switch'. In: *Proceedings of the national Academy of Sciences* 103.22 (2006), pp. 8372–8377.

[149] J. Hasty et al. 'Computational studies of gene regulatory networks: in numero molecular biology'. In: *Nature Reviews Genetics* 2.4 (2001), pp. 268–279.

[150] H. H. McAdams and A. Arkin. 'Stochastic mechanisms in gene expression'. In: *Proceedings of the National Academy of Sciences* 94.3 (1997), pp. 814–819.

[151] A. Arkin, J. Ross and H. H. McAdams. 'Stochastic kinetic analysis of developmental pathway bifurcation in phage λ-infected Escherichia coli cells'. In: *Genetics* 149.4 (1998), pp. 1633–1648.

[152] F. J. Isaacs et al. 'Prediction and measurement of an autoregulatory genetic module'. In: *Proceedings of the National Academy of Sciences* 100.13 (2003), pp. 7714–7719.

[153] J. Hasty et al. 'Noise-based switches and amplifiers for gene expression'. In: *Proceedings of the National Academy of Sciences* 97.5 (2000), pp. 2075–2080.

[154] E. M. Ozbudak et al. 'Regulation of noise in the expression of a single gene'. In: *Nature genetics* 31.1 (2002), pp. 69–73.

[155] N. Barkai and S. Leibler. 'Robustness in simple biochemical networks'. In: *Nature* 387.6636 (1997), pp. 913–917.

[156] D. Gonze, J. Halloy and A. Goldbeter. 'Robustness of circadian rhythms with respect to molecular noise'. In: *Proceedings of the National Academy of Sciences* 99.2 (2002), pp. 673–678.

[157] J. Puchaka and A. M. Kierzek. 'Bridging the gap between stochastic and deterministic regimes in the kinetic simulations of the biochemical reaction networks'. In: *Biophysical journal* 86.3 (2004), pp. 1357–1372.

[158] K. Burrage, T. Tian and P. Burrage. 'A multi-scaled approach for simulating chemical reaction systems'. In: *Progress in biophysics and molecular biology* 85.2-3 (2004), pp. 217–234.

[159] T. Akutsu, S. Miyano and S. Kuhara. 'Inferring qualitative relations in genetic networks and metabolic pathways'. In: *Bioinformatics* 16.8 (2000), pp. 727–734.

[160] L. Mao and H. Resat. 'Probabilistic representation of gene regulatory networks'. In: *Bioinformatics* 20.14 (2004), pp. 2258–2269.

[161] S. H. Strogatz. *Nonlinear dynamics and chaos with student solutions manual: With applications to physics, biology, chemistry, and engineering*. CRC press, 2018.

[162] J. Narula, M. Fujita and O. A. Igoshin. 'Functional requirements of cellular differentiation: lessons from Bacillus subtilis'. In: *Current opinion in microbiology* 34 (2016), pp. 38–46.

[163] L. Wang et al. 'Bistable switches control memory and plasticity in cellular differentiation'. In: *Proceedings of the National Academy of Sciences* 106.16 (2009), pp. 6638–6643.

[164] D. Jia et al. 'Operating principles of tristable circuits regulating cellular differentiation'. In: *Physical biology* 14.3 (2017), p. 035007.

[165] C. Gardiner, P. Zoller and P. Zoller. *Quantum noise: a handbook of Markovian and non-Markovian quantum stochastic methods with applications to quantum optics*. Springer Science & Business Media, 2004.

[166] F. Crick. 'Central dogma of molecular biology'. In: *Nature* 227.5258 (1970), pp. 561–563.

[167] N. G. Van Kampen. *Stochastic processes in physics and chemistry*. Vol. 1. Elsevier, 1992.

[168] N. Rosenfeld et al. 'Gene regulation at the single-cell level'. In: *science* 307.5717 (2005), pp. 1962–1965.

[169] H. Kitano. 'Biological robustness'. In: *Nature Reviews Genetics* 5.11 (2004), pp. 826–837.

[170] A. Skupin and M. Falcke. 'Statistical properties and information content of calcium oscillations'. In: *Genome Informatics* 18 (2007), pp. 44–53.

[171] M. Voliotis et al. 'Information transfer by leaky, heterogeneous, protein kinase signaling systems'. In: *Proceedings of the National Academy of Sciences* 111.3 (2014), E326–E333.

[172] L. Boltzmann. *Vorlesungen über gastheorie*. Vol. 1. JA Barth (A. Meiner), 1910.

[173] L. Boltzmann. 'Ueber eine von Hrn. Bartoli entdeckte Beziehung der Wärmestrahlung zum zweiten Hauptsatze'. In: *Annalen der Physik* 258.5 (1884), pp. 68–94.

[174] P. Ehrenfest and T. Ehrenfest. *The conceptual foundations of the statistical approach in mechanics.* Courier Corporation, 1990.

[175] G. Gallavotti. 'Ergodicity: a historical perspective. Equilibrium and Nonequilibrium'. In: *The European Physical Journal H* 41.3 (2016), pp. 181–259.

[176] J. Von Plato. 'Boltzmann's ergodic hypothesis'. In: *Archive for History of Exact Sciences* 42.1 (1991), pp. 71–89.

[177] L. Boltzmann. 'Zur Priorität der Auffindung der Beziehung zwischen dem zweiten Hauptsatze der mechanischen Wärmetheorie und dem Principe der kleinsten Wirkung'. In: *Annalen der physik* 219.6 (1871), pp. 211–230.

[178] L. Boltzmann. *Studien über das Gleichgewicht der lebendigen Kraft zwischen bewegten materiellen Punkten: vorgelegt in der Sitzung am 8. October 1868.* K. und k. Hof-und Staatsdr., 1868.

[179] L. Boltzmann. *Über die Beziehung zwischen dem zweiten Hauptsatze des mechanischen Wärmetheorie und der Wahrscheinlichkeitsrechnung, respective den Sätzen über das Wärmegleichgewicht.* Kk Hof-und Staatsdruckerei, 1877.

[180] L. Larue et al. 'A role for cadherins in tissue formation'. In: *Development* 122.10 (1996), pp. 3185–3194.

[181] J. Wetherington, G. Serrano and R. Dingledine. 'Astrocytes in the epileptic brain'. In: *Neuron* 58.2 (2008), pp. 168–178.

[182] J. W. Swann et al. 'Spine loss and other dendritic abnormalities in epilepsy'. In: *Hippocampus* 10.5 (2000), pp. 617–625.

[183] Y. Wu et al. 'Microglia: dynamic mediators of synapse development and plasticity'. In: *Trends in immunology* 36.10 (2015), pp. 605–613.

[184] L. J. Lawson et al. 'Heterogeneity in the distribution and morphology of microglia in the normal adult mouse brain'. In: *Neuroscience* 39.1 (1990), pp. 151–170.

[185] A. Colman-Lerner et al. 'Regulated cell-to-cell variation in a cell-fate decision system'. In: *Nature* 437.7059 (2005), pp. 699–706.

[186] M. Ogawa. 'Differentiation and proliferation of hematopoietic stem cells'. In: (1993).

[187] R. Bargaje et al. 'Cell population structure prior to bifurcation predicts efficiency of directed differentiation in human induced pluripotent cells'. In: *Proceedings of the National Academy of Sciences* 114.9 (2017), pp. 2271–2276.

[188] J. Stelling et al. 'Robustness of cellular functions'. In: *Cell* 118.6 (2004), pp. 675–685.