From the Department of Medical Biochemistry and Biophysics
Karolinska Institutet, Stockholm, Sweden

LINEAGES AND MOLECULAR HETEROGENEITY IN THE DEVELOPING NERVOUS SYSTEM

Gioele La Manno

Karolinska Institutet

Stockholm 2018

Cover art - Front cover: *Developmental lineages as a flow in gene expression space*.
Back cover: *Waddington's epigenetic landscape meets RNA velocity*.

All previously published papers were reproduced with permission from the publisher.
Published by Karolinska Institutet.
Printed by E-Print AB 2018
© Gioele La Manno, 2018
ISBN 978-91-7831-155-2

LINEAGES AND MOLECULAR HETEROGENEITY IN THE DEVELOPING NERVOUS SYSTEM

THESIS FOR DOCTORAL DEGREE (Ph.D.)

By

Gioele La Manno

Public Defense on **Friday, October 26, 2018, 10:00**
At **Biomedicum Big Lecture Hall**, Solnavägen 9, Karolinska Institutet

Principal Supervisor:
Prof. Sten Linnarsson
Karolinska Institutet
Department of Medical Biochemistry and Biophysics
Division of Molecular Neurobiology

Co-supervisors:
Prof. Ernest Arenas
Karolinska Institutet
Department of Medical Biochemistry and Biophysics
Division of of Molecular Neurobiology

Prof. Jussi Taipale
Karolinska Institutet
Department of Medical Biochemistry and Biophysics
Division of Functional Genomics

Opponent:
John Marioni PhD
University of Cambridge
Wellcome Trust Sanger Institute
and Cancer Research UK Cambridge Institute

Examination Board:
Prof. Jonas Muhr
Karolinska Institutet
Department of Cell and Molecular Biology

Ana Teixeira PhD
Karolinska Institutet
Department of Molecular Biochemistry and Biophysics
Division of Biomaterials

Prof. Per Lötstedt
Uppsala University
Department of Information Technology
Division of Scientific Computing

To Marina

To my parents

In the fields of observation chance favors only the prepared mind.

Louis Pasteur

ABSTRACT

Information in the genome unfolds through a dynamic process leading to the molecular and anatomical organization of a physiologically functional organism. The nervous system is the most diverse and intricate architecture generated by this process. It is composed of hundreds of millions of cells of hundreds of different cell types, whose connectivity and interactions are the physiological underpinnings of our capacity to respond to stimuli, our ability to learn and our cognitive capabilities.

In this thesis, I explore the formation of tissues in the nervous system during embryonic development. In particular, I focus on changes in molecular composition that lead progenitor cells to generate a complex mix of cell types. The specific aim of this work is to address the lack of complete and systematic knowledge of the heterogeneity of neural tissues and to describe the progression of a cell through different molecular states. To achieve this, I took advantage of the new opportunities offered by single-cell expression profiling technologies to gain a holistic view of a developing tissue.

To contextualize the work, I review the relevant literature and conceptual framework. Starting with a historical perspective, I discuss the concept of cell type and how it relates to developmental dynamics and evolution. I then review different aspects of developmental neuroscience, starting with general principles and then focusing on the main areas of interest: the ventral midbrain, the sympathetic nervous system, and postnatal development. Then the technological advances instrumental for this thesis are reviewed, with a focus on analysis methods for single-cell RNA sequencing. Finally, I discuss the relationship between lineages and gene regulation, and I introduce the reader to the idea of a global time derivative of gene expression through traditional systems biology modeling.

Then I present the results of three different studies.

In **paper I**, we used single-cell RNA sequencing to describe the cell-type heterogeneity of sympathetic ganglia. We found seven distinct kinds of neurons, where only two had been previously described. Using lineage tracing, we shed light on the developmental origin of the new types. We linked their molecular profile to function and described how they innervate the erector muscles.

Paper II describes the embryonic development of the ventral midbrain at the single-cell level. We characterized human and mouse embryonic tissues, identifying cell types and their homologies. We found an uncharacterized heterogeneity among radial glial cells and gained new insight into the timing of dopaminergic neurons specification. Finally, we presented a data-driven strategy to assess the quality of *in vitro* differentiation protocols.

In **paper III** we addressed the major limitation of studying development with single-cell RNA sequencing: the absence of a temporal dimension. We described an analysis framework that uses the ratio of spliced to unspliced RNA abundance to estimate the time derivative of gene expression. The method was used to predict the future molecular states of cells and to determine their fate bias.

In these studies, we produced a rich description of tissue heterogeneity and answered different biological questions. The results were achieved by harnessing the information contained in the data through analysis approaches inspired by developmental or physical principles. In summary, this thesis provides new insight into several aspects of mammalian nervous-system development, and it presents analytical approaches that I predict will inspire future investigation of the developmental dynamics of single-cells.

LIST OF SCIENTIFIC PAPERS

I. *Alessandro Furlan,* ***Gioele La Manno****, Moritz Lübke, Martin Häring, Hind Abdo, Hannah Hochgerner, Jussi Kupari, Dmitry Usoskin, Matti S Airaksinen, Guillermo Oliver, Sten Linnarsson, Patrik Ernfors*

Visceral motor neuron diversity delineates a cellular basis for nipple-and pilo-erection muscle control.

Nature Neuroscience 2016 October; 19(10), 1331-1340

II. ***Gioele La Manno*****, Daniel Gyllborg*, Simone Codeluppi, Kaneyasu Nishimura, Carmen Salto, Amit Zeisel, Lars E Borm, Simon RW Stott, Enrique M Toledo, J Carlos Villaescusa, Peter Lönnerberg, Jesper Ryge, Roger A Barker, Ernest Arenas, Sten Linnarsson*

Molecular diversity of midbrain development in mouse, human, and stem cells.

Cell 2016 October; 167(2) 566-580

III. ***Gioele La Manno****, Ruslan Soldatov, Hannah Hochgerner, Amit Zeisel, Viktor Petukhov, Maria Kastriti, Peter Lonnerberg, Alessandro Furlan, Jean Fan, Zehua Liu, David van Bruggen, Jimin Guo, Xiaoling He, Roger Barker, Erik Sundstrom, Goncalo Castelo-Branco, Igor Adameyko, Patrick Cramer, Sten Linnarsson, Peter Kharchenko*

RNA velocity of single cells

Nature 2018 August; 560(7719) 494-498

Pubblications not included in this thesis

*Codeluppi S, Borm LE, Zeisel A, **La Manno G**, van Lunteren JA, Svensson CI, Linnarsson S*
Spatial organization of the somatosensory cortex revealed by cyclic smFISH.
***Nature Methods**. In press.*

*Joost S, Jacob T, Sun X, Annusver K, **La Manno G**, Sur I, Kasper M*
Single-cell transcriptomics of traced epidermal and hair follicle stem cells reveals rapid adaptations during wound healing.
***Cell Reports.** In press.*

*Zeisel A, Hochgerner H, Lonnerberg P, Johnsson A, Memic F, van der Zwan J, Haring M, Braun E, Borm L, **La Manno G**, Codeluppi S, Furlan A, Skene N, Harris KD, Hjerling Leffler J, Arenas E, Ernfors P, Marklund U, Linnarsson S.*
Molecular architecture of the mouse nervous system.
***Cell**. 2018 Aug 9.*

*Häring M, Zeisel A, Hochgerner H, Rinwa P, Jakobsson JET, Lönnerberg P, **La Manno G**, Sharma N, Borgius C, Kiehn O, Lagerström MC, Linnarsson S*, Ernfors P*.*
Neuronal atlas of the dorsal horn defines its architecture and links sensory input to transcriptional cell types.
***Nature Neuroscience**. 2018 April 23.*

*Hochgerner H, Lönnerberg P, Hodge R, Mikes J, Heskol A, Hubschle H, Lin P, Picelli S, **La Manno G**, Ratz M, Dunne J, Husain S, Lein E, Srinivasan M, Zeisel A, Linnarsson S.*
STRT-seq-2i: dual-index 5′ single cell and nucleus RNA-seq on an addressable microwell array.
***Scientific reports**. 2017 Nov 27.*

*Toledo EM, **La Manno G**, Rivetti di Val Cervo P, Gyllborg D, Islam S, Villaescusa S, Linnarsson S, Arenas E*
Molecular analysis of the midbrain dopaminergic niche during neurogenesis.
***bioRxiv**. 2017 Jun 26.*

*Rivetti di Val Cervo P, Romanov RA, Spigolon G, Masini D, Martín-Montañez E, Toledo EM, **La Manno G**, Feyder M, Pifl C, Ng YH, Sánchez SP, Linnarsson S, Wernig M, Harkany T, Fisone G, Arenas E.*
Induction of functional dopamine neurons from human astrocytes in vitro and mouse astrocytes in a Parkinson's model.
***Nature Biotechnology**. 2017 Oct 10.*

*Joost S, Zeisel A, Jacob T, Sun X, **La Manno G**, Lönnerberg P, Linnarsson S, Kasper M*
Single-Cell Transcriptomics Reveals that Differentiation and Spatial Signatures Shape Epidermal and Hair Follicle Heterogeneity.
***Cell Systems**. 2016 Sep 28.*

*Marques S, Zeisel A, Codeluppi S, van Bruggen D, Mendanha Falcão A, Xiao L, Li H, Häring M, Hochgerner H, Romanov RA, Gyllborg D, Muñoz-Manchado AB, **La Manno G**, Lönnerberg P, Floriddia EM, Rezayee F, Ernfors P, Arenas E, Hjerling-Leffler J, Harkany T, Richardson WD, Linnarsson S, Castelo-Branco G*
Oligodendrocyte heterogeneity in the mouse juvenile and adult central nervous system.
***Science**. 2016 June 10.*

*Zeisel A, Muñoz-Manchado AB, Codeluppi S, Lönnerberg P, **La Manno G**, Juréus A, Marques S, Munguba H, He L, Betsholtz C, Rolny C, Castelo-Branco G, Hjerling-Leffler J, Linnarsson S.*
Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq.
***Science** 2015 March 6.*

*Islam S, Zeisel A, Joost S, **La Manno G**, Zajac P, Kasper M, Lönnerberg P, Linnarsson S.*
Quantitative single-cell RNA-seq with unique molecular identifiers.
***Nature Methods** 2014 Feb.*

CONTENTS

1 Cell Types: Historical And Modern Concept 3
- 1.1 Historical introduction 3
- 1.2 Characterization of neural cell types 5
- 1.3 An evolutionary definition 6
- 1.4 Intermediate cell states 9
- 1.5 Remarks on the terminology 11

2 Development And Patterning of The Nervous System 13
- 2.1 General principles of central nervous system development 13
- 2.2 Postnatal development and neurogenesis 14
- 2.3 Ventral midbrain development and Parkinson's disease 16
- 2.4 Origin of the peripheral nervous system 18
- 2.5 Development of the sympathetic ganglia 18

3 Single-Cell RNA Sequencing 21
- 3.1 The evolution of the technique 21
- 3.2 Analysis of single-cell data 22
- 3.3 Data analysis for cell type discovery 25
- 3.4 Extension of the analysis to dynamical systems 27

4 The Cybernetics Of Development 31
- 4.1 Timescales and rhythm 31
- 4.2 Lineages 32
  - 4.2.1 Lineage as mitotic kinship 32
  - 4.2.2 Lineage tracing 33
  - 4.2.3 Lineage as the progression through molecular states 34
  - 4.2.4 Combined lineage-phenotype analysis 35
- 4.3 Mathematical modeling of development 36
- 4.4 Relation Between Dynamic and Evolutionary Perspective 39

5 Results 41
- 5.1 Paper I - Visceral motor neuron diversity delineates a cellular basis for nipple-and pilo-erection muscle control 41
- 5.2 Paper II - Molecular diversity of midbrain development in mouse, human, and stem cells 42
- 5.3 Paper III – RNA velocity of single cells 43

6 Conclusion And Perspectives 44

7 Acknowledgements 47

8 References 51

LIST OF ABBREVIATIONS

BMP	Bone morphogenetic protein
CA	Cornu Ammonis
CNS	Central nervous system
CNV	Copy number variation
ICA	Independent component analysis
iPSC	Induced pluripotent stem cells
kNN	k-Nearest neighbors
NEM	Nipple-erection muscle
OPC	Oligodendrocyte precursor cell
PCR	Polymerase chain reaction
PEM	Pilo-erection muscle
RNA-seq	RNA sequencing
scRNA-seq	Single-cell RNA sequencing
smFISH	Single molecule fluorescent in situ hybridization
t-SNE	t-Distributed stochastic neighbor embedding
UMI	Unique molecular identifier

“Many years ago, when comparing, and seeing others compare, the birds from the separate islands of the Galapagos Archipelago […] *I was much struck how entirely vague and arbitrary is the distinction between species and varieties.* […]
Certainly no clear line of demarcation has as yet been drawn between species and sub-species […] *or, again, between sub-species and well-marked varieties, or between lesser varieties and individual differences. These differences blend into each other in an insensible series; and a series impresses the mind with the idea of an actual passage.”*

Charles Darwin - The origin of species, 1860

1 CELL TYPES: HISTORICAL AND MODERN CONCEPT

1.1 HISTORICAL INTRODUCTION

With the establishment of cell theory in the 19th century, (Remak, 1852; Schleiden et al., 1847; Schwann, 1847) biologists became aware that cells were the building blocks of every life form. This awareness shifted the attention of scientists to the great diversity of cell organization, function and specialization. No two cells are phenotypically identical. Cellular heterogeneity started to be accurately described, and boundaries were defined both between different taxonomic units (e.g. the cytological basis of the difference between domains (Chatton, 1938)) and within cells of the same multicellular organism (Ramón Cajal, 1904). It was in the first half of the 20th century that the growing number of observations and theories linking structure to physiology shaped histology into a scientific discipline of its right (Musumeci, 2014). The development of histology was strongly linked to the tools, staining and visualization techniques that were available at a particular time. Therefore, it is not surprising that the characterization of cell types started as purely morphological, with the definition of the so-called morphotypes (Hall, 2007). As they became available, new tools, such as electron microscopy, modern imaging techniques and functional assays, started to add cytoanatomical, molecular, and biochemical perspective to the morphological definition, helping to refine cell type classifications (Somogyi and Klausberger, 2005; Valentine, 2002).

Histology, as the science of tissues (from the Greek: histos: tissue, logia: branch of learning), describes cell types in reference to their tissue context and, in a sense, is limited by this property. For example, similar cells in different locations might be described as distinct, when they are, instead, related. An example to illustrate this point is the classification of tissue-resident macrophages, a group of cell types that are named differently depending on the tissues where they are found: Langerhans cells in the liver, Kupffer cells in the pancreas, alveolar macrophages in the lung, red pulp macrophages in the spleen, and microglia in the central nervous system (CNS). The nomenclature does not indicate that these cells belong to the same mononuclear phagocyte system, nor does their phenotypic description suggest that microglia constitute an ontogenetically distinct population (Ginhoux et al., 2010; Prinz and Priller, 2014).

Despite the fact that the question is almost two centuries old, there is no universally accepted definition of what a cell type is (Clevers et al., 2017). As a consequence, a catalog of all mammalian cell types is lacking, and projections on their putative number by different scholars sometimes differ by more than an order of magnitude. One of the most notable attempts to review the classification of human cell types is by Vickaryous and Hall. Their reordering of previous knowledge is a useful starting point for a complete classification but does not offer a resolution to the above-mentioned debates (Vickaryous and Hall, 2006).

Authorities of the field seem to agree that different morphological, cytoanatomical, molecular and biochemical properties can be used to define distinct, non-overlapping classes of cells (Alberts et al., 2014; Clevers et al., 2017; Valentine, 2002). This statement implies that a phenotypic definition of cell type makes sense and that the variation found between two members of the same cell type can be attributed to the stochasticity governing molecular dynamics. While intuitive, this fact has been difficult to prove in the past and only with recent technologies has it become possible to produce convincing evidence that cells naturally cluster in well defined phenotypic subspaces (Bendall et al., 2011; Macosko et al., 2015a; Zeisel et al., 2015).

This result has been recently achieved using methods such as mass cytometry and single-cell RNA sequencing that enable the collection of high-dimensional single-cell phenotypic data. Furthermore, these high-throughput studies have shown that while a phenotypic gap exists between mature cell types, a continuum of intermediate characters exists between immature and mature cell types (Alberts et al., 2014; Bendall et al., 2014; Trapnell, 2015). High-throughput technologies have made unbiased classification possible and rendered previous classification susceptible to updates and reinterpretations (Furlan et al., 2016; Usoskin et al., 2014). This process has even involved disciplines like immunology, where surface marker-based cell-type classification appeared robust and well delineated (Giladi and Amit, 2018; Jaitin et al., 2016; Paul et al., 2015).

Although single-cell RNA sequencing aims to define cell types in an unbiased way, the fact that these types are defined from a particular phenotypic readout might be seen as just another bias (Shapiro et al., 2013). Another critique arises when relating molecular phenotyping to the concept of function. In particular, it becomes difficult to decide whether a cell should still be regarded as the same cell type after a modification that causes its functional impairment. An example is the introduction of a missense mutation that disrupts the function of a protein essential for the activity of the cell, such as a receptor or an enzyme, but leaves the cell's molecular composition otherwise unaffected. Does the cell so modified belong to the same cell type? This and other extreme scenarios can be easily imagined and used to challenge both the phenotypic and the functional concepts of cell type. In this context, it is desirable to anchor the concept of cell type to a solid theory to enable it to resist degenerated cases. Before continuing on this line of thought, and discussing how an evolutionary perspective centered on gene regulation can confer robustness to this concept, I give a short overview of the scientific efforts directed at characterizing the subset of cell types that are the focus of this thesis: the cell types of the nervous system.

1.2 CHARACTERIZATION OF NEURAL CELL TYPES

The mammalian brain consists of hundreds of regions that are distinct in structure, cell density and composition. The complexity of electrophysiological responses and behaviors that are generated by the brain requires a large number of specialized cell types. These types include not only neurons but also a variety of glial types that ensure trophic support and maintenance of the homeostatic conditions (Aloisi, 2001; Magistretti, 2011; Tsacopoulos et al., 1997).

The relation between the variety of neuronal types and the computational capabilities of the nervous system has rendered the identification and characterization of neuronal types as a prerequisite to understanding the brain (Ramón Cajal, 1904; Sugino et al., 2006). This goal has been pursued using both tissue-agnostic classification criteria like location, morphology and molecular markers, and more field-specific criteria such as anatomical projections, target specificity, synaptic temporal dynamics and electrophysiological activity (Klausberger and Somogyi, 2008; Molyneaux et al., 2007). The classification has reached various degrees of accuracy depending on the area of the brain. More detailed information is available for brain regions that are traditionally well studied due to their involvement in higher cognitive functions (e.g. the telencephalon), sensory input (e.g. retina) or their relevance in pathologies (e.g. substantia nigra and ventral tegmental area) (Fishell and Rudy, 2011; Fu et al., 2012; Poulin et al., 2014; Somogyi and Klausberger, 2005).

For example, the CA1 (CA: cornu ammonis) area of the hippocampus, owing to its relative simplicity, has recently been described to satisfactory detail in terms of neuronal diversity, connectivity and activity. In this region, three kinds of pyramidal cells are responsible for encoding spatial representation and other episodic memories (O'Keefe, 1976; Quiroga et al., 2005). These neurons contact at least ten extra hippocampal brain areas, and a vast collection of GABAergic interneurons supports their function. CA1 interneurons are classified into basket, axo-axonic, bistratified and oriens–lacunosum moleculare interneurons by the subcellular domain they innervate, or alternatively into cholecystokinin, parvalbumin, calbindin expressing cells and cholinergic interneurons on the basis of their molecular signature (DeFelipe et al., 2013; Klausberger and Somogyi, 2008).

Although regions such as the CA1 have been characterized with an excellent level of accuracy, a more generalized characterization including all areas of the brain is still missing. In this context, the Allen Institute of Brain Science made an impressive systematic effort to resolve extensive regional heterogeneity with automated *in situ* hybridization experiments (Lein et al., 2007). This collection of gene-expression profiles is organized in an atlas that has become a fundamental resource for neuroscientists. The resource includes a growing number of datasets from the mouse developing brain and the human brain. The fine resolution of the resource (cellular but not single cell) provides meaningful local correlations between gene

expression and cellular phenotype in different areas of the CNS. However, in the Allen Brain atlas, every gene is detected in an independent sample. Therefore it is not possible to gain information on whether a gene co-localizes in the same cell and ultimately this impairs cell type discovery from this kind of data.

On the other hand, single-cell RNA-seq provides data of the appropriate resolution to undertake cell-type discovery. We have shown, for the first time, that this approach can be used successfully to explore the heterogeneity of the nervous system (Usoskin et al., 2014; Zeisel et al., 2015). Since then, the approach has been used to molecularly map neurons in several areas of the brain including the cortex, the hippocampus, the hypothalamus and the sympathetic ganglia but also to study other lineages in the central nervous system such as the oligodendrocytes and microglia (Furlan et al., 2016; Goldmann et al., 2016; Marques et al., 2016; Romanov et al., 2016; Tasic et al., 2016).

Notably, two very recent contributions by the McCarroll and Linnarsson labs constitute essential steps towards the definition of a complete cell-type atlas of the nervous system (Saunders et al., 2018; Zeisel et al., 2018). In Zeisel et al. 500.000 single cells from different areas of the central and peripheral nervous system were profiled and 265 cell types defined. The data-driven hierarchical taxonomy that resulted from this analysis has tremendous value as a resource. Using this taxonomy as a starting point, the overall architecture of the nervous system can be explored and discussed, and the major principles of its organization extracted.

1.3 AN EVOLUTIONARY DEFINITION

Following a meeting held at the Santa Fe Institute in 2016, a definition of cell types as evolutionary units was formulated. The result of the discussions are collected and organized in a consistent theory in Arendt et al., 2016. Their definition reads "A [cell type is] set of cells in an organism that change in evolution together, partially independent of other cells, and are evolutionarily more closely related to each other than to other cells". This description has the advantage of being technique agnostic and generalizable to several biological disciplines and, most importantly, links the idea of cell type to evolution, the central driving force of biology. The definition implies that some genomic information exists to be exclusively accessible to one set of cells and not to others. This genomic information consists of regulatory elements, enhancers and gene products that cooperate to generate specific patterns of expression and biochemical compositions.

The Santa Fe working group established the fundamental concepts and cornerstone nomenclature necessary to discuss cell types with an evolutionary perspective. They stated that independent regulation is necessary for the evolution of a new gene-expression profile (Figure 1) and introduced the idea of a *core regulatory complex* through which this is achieved. A *core regulatory complex* consists of a particular combination of transcription

factors ("terminal selector") that interact with regulatory elements and regulate gene expression. When a new core regulatory complex arises, it creates an opportunity for the independent regulation of one or more genes modules and, therefore, for the genetic individuation of a new cell type.

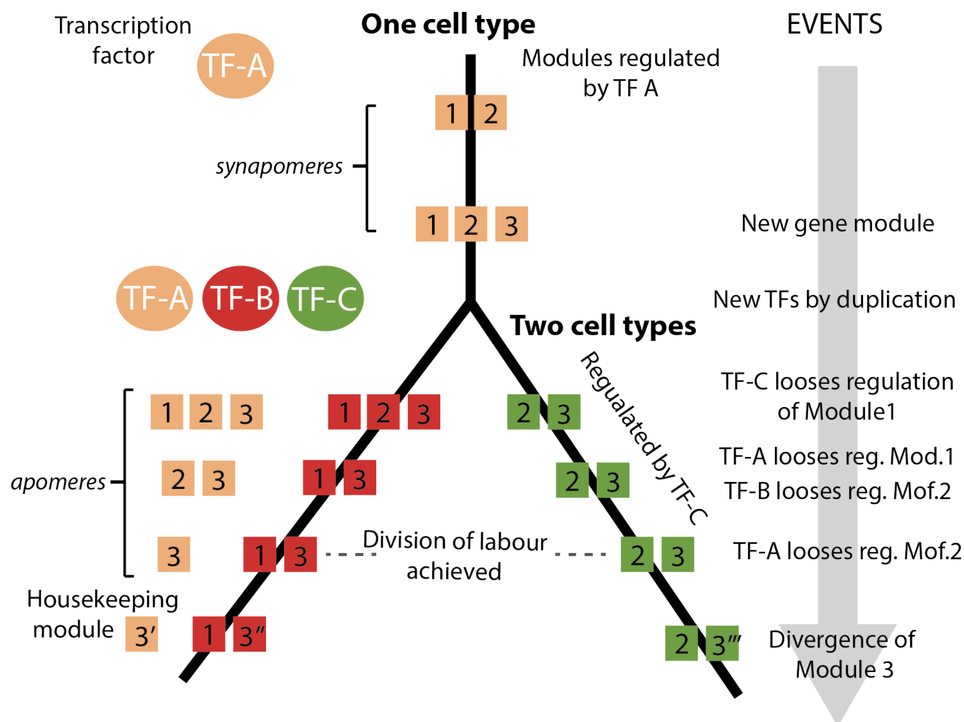


Figure 1 – A simple model of the evolution of sister cell types by genetic individuation

Another important set of terms was introduced to describe new modules of gene regulation that arise through the evolutionary process (Figure 1). *Apomeres* (coined by analogy to the term *apomorphy*) are modules of gene expression that result from division of labor, molecular divergence or neofunctionalization. In contrast, *synapomeres* are ancestral modules that are shared by sister cell types.

Analogous to the concept of anatomical homology postulated by Owen as "the same organ in different animals under every variety of form and function", cell-type homology requires cells to share the same regulation plan rather than function (Owen, 1848). The evolutionary definition abandons a purely phenotypic framework and instead embraces a description centered on gene regulation, focusing on the events that determine genetic individuation of cell types (Figure 2). When data on these events are available, the new focus can correctly disambiguate some complicated conundrums, for example, interpreting the relation between ciliated photoreceptors with different functions (Brunet et al., 2016). However, the evolutionary definition is challenging to transform into systematic operational criteria for defining cell types from data.

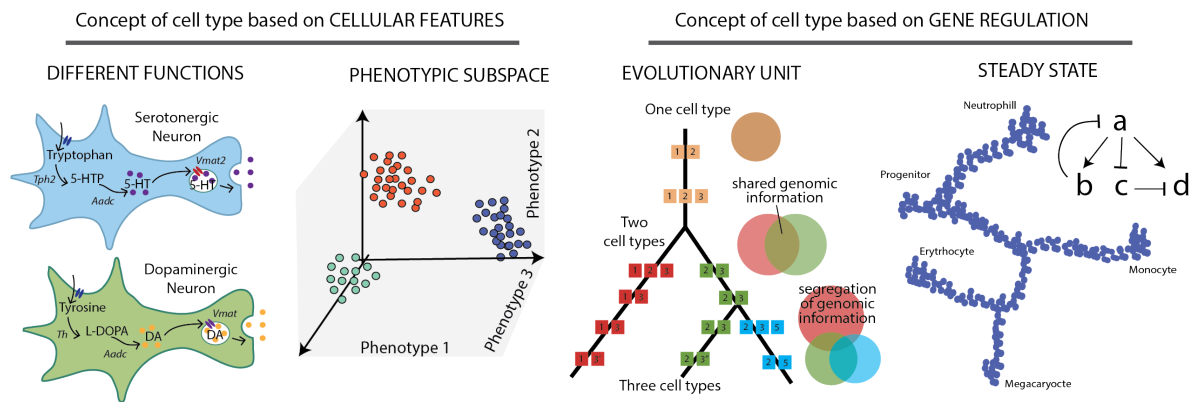


Figure 2 – Different definitions of cell type focus on different aspects of the concept

In a recent review, Marioni and Arendt discuss this operational challenge. They summarize the efforts of the scientific community to better understand of the evolution of cell types (Marioni and Arendt, 2017). Among the tools available at the moment, they recognize single-cell transcriptomics as the best candidate to start tackling the difficult task. The choice is circumstantial; among recent techniques single-cell transcriptomics provides the most significant amount of information on the regulatory processes. The idea is that, if the physical events related to regulation cannot be measured directly, one can use transcriptomics to find sets of cell type-specific genes that constitute candidates for core regulatory complexes.

In Paper II of this thesis, we present an initial attempt to move the phenotypic definition of cell types closer to the evolutionary perspective by establishing homology relationships between mouse and human cell states, defined on the basis of their gene-expression profiles. In the study, we calculate the correlation between cell-type expression profiles by matching gene orthologs one-to-one between the species. We thereby define *bona fide* homologous cell types as the best one-to-one cell types between species. In our study and successive pubblications, the phenotypical datasets are used in a comparative way to find cell type-specific gene expression or to reconstruct the evolution of cell types (Sebé-Pedrós et al., 2018a; Tosches et al., 2018).

Until extensive data from multiple species become available, and gene orthology relations are well defined and annotated, the evolutionary framework will be difficult to delineate accurately. The difficulty arises not only from the limited amount of data and gene annotation but also from the absence of a quantitative framework to guide analytical efforts. The definition of a framework for analysis presents several challenges and requires answering important questions related to modeling and data analysis. For example, the definition of distance between two single-cell transcriptomes must account for orthogonal processes like cell cycle and circadian rhythm, and mappings from the gene-expression space of one species to another must be defined.

It is my opinion that a principled framework, with the cell as a biochemical dynamical system at its core, is crucial to find an adequately rigorous solution to these problems. I will discuss this idea further in the last chapter, where I present cell types from a systems biology perspective (Figure 2). The discussion will be largely conceptual, however, because the techniques currently available are unable to estimate the millions of biochemical parameters necessary to describe such a dynamical system.

At this point in time, it makes sense to consider transcriptional states as *bona fide* cell types, not only because of practical considerations like the availability of single-cell RNA-seq and other high-throughput single-cell techniques (discussed later in this thesis), but also for at least two other reasons. The first is that any transcriptional steady state should always depend on the same set of regulatory elements and regulators, even if the state is attainable through different developmental paths. The second is that an intrinsic evolutionary dimension exists even when comparing cell types from a single species. The different cell types are themselves evolutionarily related (sister cells types) and can be thought of as leaves on a phylogenetic tree (Arendt, 2008).

1.4 INTERMEDIATE CELL STATES

It is reasonable to object that a description of cell types as discrete units is too restrictive. This objection is supported by evidence for several continua in biological systems. The concept of an intermediate cell state is often used in a poorly defined fashion and vaguely refers to the existence of a progression between two extremes. In developmental biology, the term is often used to describe two conceptually different scenarios: a time-dependent intermediate, such as a cell differentiating to a more mature state; or a input-dependent intermediate, such as a cell responding to a morphogenetic gradient in a concentration-dependent fashion. The latter should be though in a time-independent way, focusing on the fact that a different steady state is reached and maintained for each given value of the input.

Before moving to a more rigorous reasoning let us consider an example where intermediate cell states can be easily identified: the zonation of the liver. The liver is composed of hexagonal lobules consisting of, approximately, fifteen concentric layers of hepatocytes, organized around a central vein and with portal triads at the vertexes. Hepatocytes along the lobule axis respond to the porto-centrally directed blood flow and an oppositely oriented WNT gradient with a substantial change in their gene-expression profiles (Jungermann and Keitzmann, 1996). The graded response is functionally significant because it spatially segregates the steps of different enzymatic cascades (Gebhardt, 1992).

Zonated hepatocytes are a good example of intermediate states in reference to a concentration gradient. A recent transcriptomic analysis shows that over 2000 liver genes are zonated, vary non-linearly in space and can peak in mid-lobule layers (Halpern et al., 2017). This non-

linearity is the result of a non-trivial regulation of gene expression. It hints at the fact that the gene expression of an intermediate (defined in relation to an external variable) is not, in general, a weighted average of the two extremes.

Reasoning in terms of gene-expression space, one might conceive that an intermediate state is a point *C* that sits on a line that connects two extremes *A* and *B*. In other words, that *C* is an affine combination of the kind:

$$C = \alpha A + (1 - \alpha)B \quad \text{with} \quad 0 < \alpha < 1$$

However, in a simple simulation where the concentration of a regulator *r* affects the expression of both *x* and *y*, the intermediate steady states generated at different concentrations of *r* do not line up connecting the extremes *A* and *B,* but instead form a curve (Figure 3).

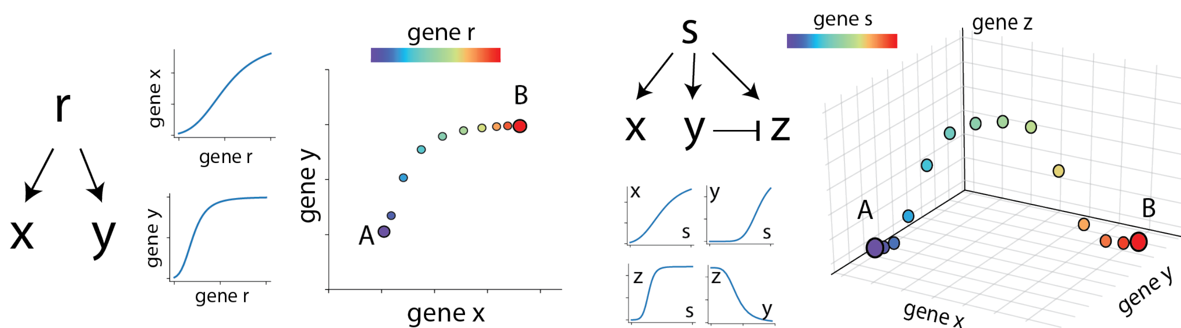


Figure 3 – Intermediate states with respect to a regulatory network and an input varaible.

With this example in mind, one might be tempted to think of an intermediate state, more generally, as any possible linear combination of the two extreme states:

$$C = \alpha_1 A + \alpha_2 B$$

This model accommodates any activation function (even non-monotonic) and allows any point on the plane spanned by the two vectors A and B. However, an intermediate is not constrained to sit in such a plane as it can be seen in a trivial situation involving 3 genes, with gene *s* inducing *x, y* and *z*, while *y* represses *z*. As the simulation in Figure 3 shows,,intermediate states can lie outside the plane spanned by A and B. Furthermore, note that regulation constrains the possible steady states intermediates between A and B to a lower dimensional manifold a U-shaped curve.

Therefore, the idea of an intermediate state as a mixture of extreme states is generally wrong. Instead, gene regulatory networks are responsible for complex curves in gene-expression space. It follows that discretely sampled molecular states cannot be ordered in a progression *per se* but only in relation to a given process and an input variable. Understanding regulation is, ultimately, the only way to rigorously define a state as intermediate and to determine the order of a progression.

1.5 REMARKS ON THE TERMINOLOGY

I have now briefly laid out the historical emergence of the concept of cell type and presented the controversies that naturally arise in the attempt to accurately define the concept. To resolve the controversies, I suggested it is convenient to adopt a perspective centered on evolution and gene regulation. This perspective is valuable but not a definitive solution, and its full potential is unlocked when combined with a dynamical view of biological processes, as we discuss later in this thesis.

In writing this thesis, I had to face the necessity of talking about cell types without the support of a compendium of cell types recognized by the scientific community. Therefore, I use the term *cell type* to*,* more casually, indicate a subset of cells that closely resemble each other transcriptionally and that can be observed at homeostatic conditions. I use the term *subtype* to instead identify cells whose distinction relies on few variables, and the term *cell state* to more generally include even cells undergoing a process and not necessarily in homeostatic or steady-state conditions.

While this choice might be perceived as a regression after the aforementioned discussion, we believe it is necessary. It is consoling to realize that even Charles Darwin, in *On the Origin of species* was forced to a similarly ambiguous choice of convenience: "*I look at the term species, as one arbitrarily given for the sake of convenience to a set of individuals closely resembling each other, and that it does not essentially differ from the term variety, which is given to less distinct and more fluctuating forms [...] also applied arbitrarily, and for mere convenience' sake*".

"The experiments which finally led to the discovery of the phenomena which are now designated as "organizer-effect" were prompted by a question which actually goes back to the beginnings of developmental mechanics, indeed to the beginnings of the history of evolution in general. How does that harmonious interlocking of separate processes come about which makes up the complete process of development? Do they go on side by side independently of each other by "self-differentiation", but from the very beginning so in equilibrium that they form the highly complicated end product of the complete organism, or is their influence on each other one of mutual stimulation, advancement or limitation?"

Hans Spemann – Nobel Lecture, *1935*

2 DEVELOPMENT AND PATTERNING OF THE NERVOUS SYSTEM

2.1 GENERAL PRINCIPLES OF CENTRAL NERVOUS SYSTEM DEVELOPMENT

The formation of a functional adult CNS involves the generation of a diverse repertoire of neurons and glial cells. These cells are generated during embryonic development starting from mitotically active progenitors through a wide range of signaling events, regulation processes and molecular mechanisms. The molecular description of the steps through which these progenitors cells generate uniquely fated neurons and glia is one of the primary goals of developmental neurobiology (Brody and Odenwald, 2005).

The CNS is formed at the end of gastrulation under the influence of SHH an inductive factor derived from the notochord (Figure 4). In a process that goes under the name of neurulation, BMP signaling initiates the folding of the neural plate to form the neural tube and instructs ectodermal cells to commit to a neural fate (Grove and Monuki, 2013). As a result as set of cell populations is established within the neuroectoderm that provides local sources of signals within the tissue. These cells constitute the so-called organizers, a term originally used in reference to the Spemann organizer. The term has been extended to refer to any group of cells that acts as a signaling center and can induce a fate change in neighboring cells (Harland and Gerhart, 1997). Early organizers, such as the floor plate, provide a fundamental but crude initial patterning of the neural tube that is later refined and later modulated by local (or secondary) organizers such as the zona limitans intrathalamica and the midbrain-hindbrain boundary (Kiecker and Lumsden, 2012).

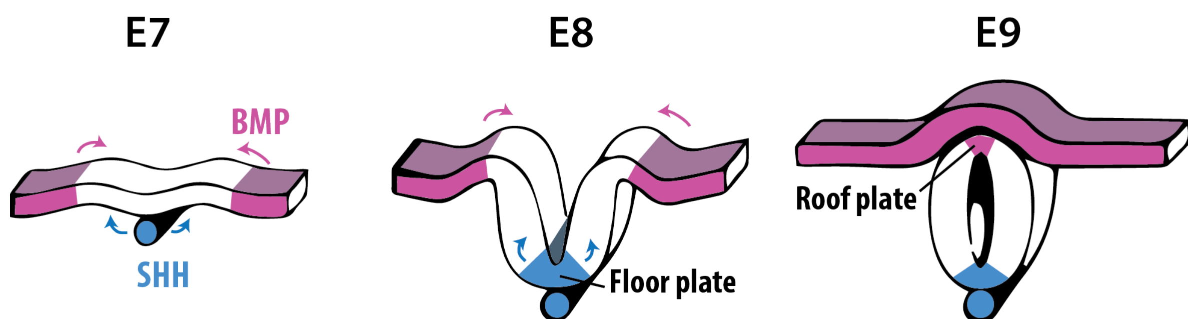


Figure 4 – Formation of the neural tube induced by the notocord and epidermal ectoderm

To achieve the great complexity of the CNS, the neural tube is eventually compartmentalized into different areas. Over time, cell-intrinsic mechanisms, proximity signals and other interactions between cells determine the identities of specific cell types in each area. A tissue where this process has been well characterized is the cerebral cortex.

In the cortex, both cell type and regional identity are specified in a stepwise fashion (Leone et al., 2008). Transcription factor gradients that encompass both progenitors and neuroblasts

(early postmitotic neuronal cells) start to impart regional identity at E11.5. Only later, close to the time of birth, does this initial patterning become more sharply restricted, defining cortical areas (O'Leary et al., 2013). Similarly, cell type identity is progressively specified and new cell types are generated from the ventricular zone, leading to the formation of cortical layers in an inside-out order.

Radial glial cells play a central role in this process: they are the main stem cells of the nervous system, which divide asymmetrically to generate intermediate progenitors and self-renew (Kriegstein and Alvarez-Buylla, 2009). They are capable of giving rise to neurons, oligodendrocytes, astrocytes and ependymal cells through successive waves of divisions (Shen et al., 2006). Their name derives from their characteristic morphology, with projections extending dorsoventrally and anchoring these cells to both the ventricular lumen and anchoring these cells to both the ventricular lumen and the basal lamina.

While the process of layer formation in the cortex is a peculiarity in the CNS, similar kinds of neural stem cells with a radial glial expression profile have been described in other areas of the brain and are generally referred to as radial glia-like cells (Anthony et al., 2004; Bonilla et al., 2008). Once neuroblasts are generated from radial glia-like cells, they differentiate to specific neuronal subtypes as instructed by different combinatorial transcription factor programs (MacDonald et al., 2013).

2.2 POSTNATAL DEVELOPMENT AND NEUROGENESIS

In mammals, several organs are functional at birth. However, in other tissues, the developmental processes cannot be considered complete until a later time. In the central nervous system, multiple different phenomena and cellular interactions take place after birth that affect the cell type composition of tissues. Changes that involve the neuronal lineage can be summarized in two main processes: the generation of new neuronal cells, termed neurogenesis, and maturation of both embryonically born and postnatally born neurons.

The terms postnatal and adult neurogenesis refer to the birth of new neurons in the mammalian brain after birth and in adult age, respectively. Despite the skepticism and dismissal that accompanied the initial findings, adult neurogenesis is now universally accepted, after the accumulation of an important body of evidence (Altman, 1963; Imayoshi et al., 2008). The phenomenon does not involve the whole brain (at least in mammals) but only specific niches that are often referred to as "neurogenic regions". The earliest evidence for a neurogenic region was the observation that cells in the subventricular zone, a layer of cells lining the lateral ventricle, divide postnatally, and their progeny migrates anteriorly towards the olfactory bulb (Altman and Das, 1966; Lois and Alarez-Buylla, 1994). This migratory route, termed the "rostral migratory stream", is estimated to contribute up to 90% of the olfactory bulb granule cells. The second neurogenic area identified is the subgranular

zone of the dentate gyrus, a proliferative niche that contributes new granule cells at early postnatal time points, generating ~1% of the total neuronal pool per day (Cameron and Mckay, 2001). As in embryonic development, these postnatal proliferative cells are radial glia-like. However, their phenotype is not identical to the embryonic counterpart, and this late radial glia-like cell has been described to display a more mature astrocyte-like character (Hochgerner et al., 2018; Kriegstein and Alvarez-Buylla, 2009; La Manno et al., 2016). Radial glial cells are the origin of the lineage tree of the hippocampus, including both the dentate gyrus granule neurons and the pyramidal cells of the CA1-3 and subiculum (Angevine, 1965; Malatesta et al., 2003). The close lineage relationship between these cells is highlighted by the fact that knocking-out *Prox1*, a transcription factor required for the formation of granule neurons, produces a switch from granule neurons to a CA3 pyramidal neuron fate. The CA fields are not generated by committed precursors; instead, field identity depends on interactions between cells and their microenvironment and is not specified earlier in the lineage (Grove et al., 1992). Mature cellular identity emerges gradually and, even after terminal fate commitment, adjustments in gene expression are required so that a mature and a fully functional phenotype can be reached.

In the nervous system, postnatal changes can be stark. For example, transcriptomic analysis of different brain areas has revealed that a tissue from early embryonic stages differs from the tissue at birth as much as the latter differs from adult tissue (Bakken et al., 2016). This difference might appear counterintuitive, since the neuronal pool has already been generated in many brain regions, and cells have projected axons contacting other regions. However, many processes reach completion only after birth, and others are peculiar to postnatal development. This fact was corroborated by a holistic transcriptomics analysis finding that distinct gene sets vary in pre- and postnatal development (Bakken et al., 2016). For example, synapse development is completed postnatally, a process that involves synchronized changes at the presynaptic and postsynaptic levels and the pruning of projections that do not reach their targets (Gonzalez-Lozano et al., 2016; Vanderhaeghen and Cheng, 2010). The subsequent integration of neurons into different kinds of networks can have critical activity-related effects on the transcriptome and phenotype. This has been observed for cortical pyramidal neurons, which acquire areal and laminar molecular phenotypes only late in postnatal development.

Finally, changes in the cell-type composition of a tissue and the corresponding microenvironment can cause transcriptional responses in neighboring cells. The generation of oligodendrocytes, whose first appearance can be dated around birth, is probably the most straightforward example of a process changing the structure of the neural tissue. Oligodendrocytes are produced by the proliferation of oligodendrocyte progenitor cells

(OPCs) during the first postnatal month, when they start myelinating the surrounding axons, profoundly changing the microstructure of the tissue (Qian et al., 2000).

2.3 VENTRAL MIDBRAIN DEVELOPMENT AND PARKINSON'S DISEASE

The ventral midbrain is a part of the brain whose development has been extensivelly studied, particularly in connection to Parkinson's disease, the second most common neurodegenerative pathology after Alzheimer's. Parkinson's disease is named after the 19th century physician, James Parkinson, that first described its symptoms: tremor, bradykinesia, rigidity and postural instability (Parkinson, 2002). Only a century later the disease was characterized histopathologically, by Frederic Lewy, and was found to be caused by the progressive death of dopaminergic neurons in the substantia nigra pars compacta (Holdorff, 2002).

Interest in the details of dopaminergic lineage development is fostered by the possibility that knowledge of this process could help to develop new therapies for Parkinson's disease. Current treatments for Parkinson's disease alleviate the symptoms but fail to address the cause of the disease. At the moment, arresting or effectively slowing down the progression of the disease is not possible. In this context, alternative therapeutic approaches that aim at the regeneration or replacement of degenerated neurons are being explored. In particular, cell-replacement therapies using human mesencephalic fetal tissue have shown promising initial results in clinical trials (Lindvall and Kokaia, 2009). The approach is currently being refined and further investigated through more extensive trials (Barker et al., 2015). However, to guarantee the safety and reproducibility required for clinical adoption, cells must be derived from standardized, easily accessible and scalable sources. To this purpose, patient-derived induced pluripotent stem cells (iPSC) or embryonic stem cells (ESC) have been envisioned as the best alternative, supported by evidence that these cells can differentiate into dopaminergic neurons (Arenas et al., 2015).

The path that leads to safer and more effective cell-replacement therapy passes through the acquisition of a more detailed picture of midbrain development. This knowledge will not only help to assess how similar *in vitro* cells are to their *in vivo* counterparts, but also to learn how to recapitulate *in vivo* differentiation.

Ventral midbrain development has been thoroughly studied in mice. After neurulation, three important organizers of the midbrain are formed: the floor plate, the dorsal midline and the midbrain-hindbrain boundary. These floor plate is generated as the result of SHH, a morphogen initially synthesized by the notochord (and later by the floor plate). The midbrain-hindbrain boundary is formed by expression of two transcription factors OTX2 (anterior) and GBX2 (posterior) (Figure 5) and, together with the dorsal midline, secretes two morphogens essential for ventral midbrain development: FGF8 and WNT1 (Nakamura, 2013).

Midbrain dopaminergic neuron development is triggered within the floor plate by the expression of Lmx1a, a target of OTX2, and the activation of the beta catenin pathway in response to WNT1 signaling (Chung et al., 2009). More laterally, where the level of SHH is low, the basal plate program is triggered instead (Figure 4) (Prakash et al., 2009).

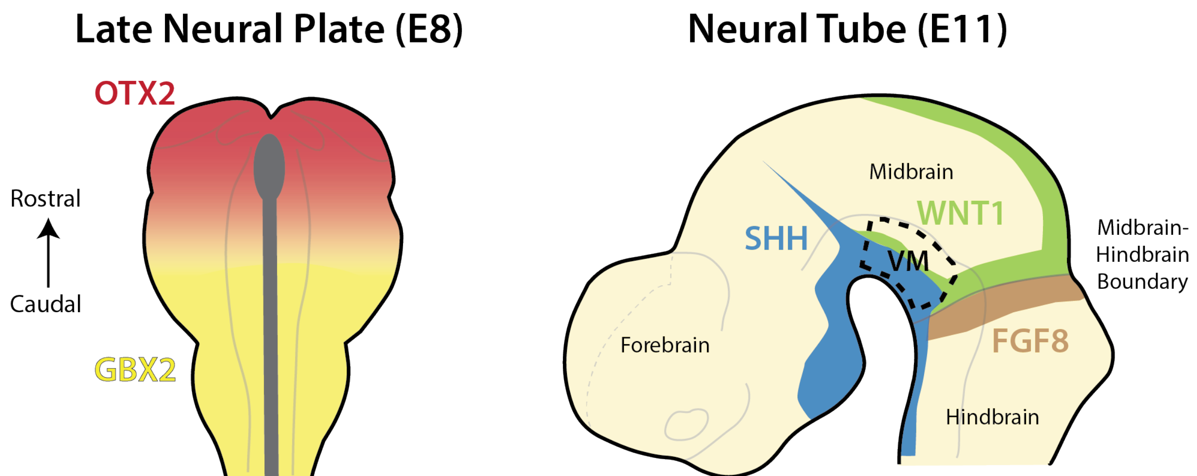


Figure 5 – Transcription factors and singaling molecules patterning the nervous system.

Further steps leading to the development of dopaminergic neurons have been characterized and involve the successive activation of key transcription factors such as Nr4a2 and Pitx3. In contrast, bifurcation events leading to the segregated dopaminergic populations of the substantia nigra and ventral tegmental area are less well understood, despite the implication for Parkinson's disease, in which substantia nigra neurons degenerate (Damier et al., 1999).

The differentiation between the substantia nigra and ventral tegmental populations has motivated efforts to count the populations of dopaminergic neurons that populate the midbrain. This work has expanded the classification from two fundamental types to more; for example, a classification based on connectivity and electrophysiological recording arrived at 13 dopaminergic populations (Roeper, 2013). Attempts to molecularly profile different populations distinguished fewer types (Chung et al., 2005). The most recent example of these attempts used single-cell real-time PCR profiling of a curated gene set to discover five dopaminergic neuron populations in the adolescent mouse (Poulin et al., 2014).

This discrepancy between the numbers of phenotypically and molecularly defined cell types is just another reminder of the necessity for systematic and unbiased molecular characterization of these types. Furthermore, in a panorama where molecular details were mainly studied in mice and chicken embryos, profiling of human development could have a critical translational impact. The knowledge of similarities and peculiarities might turn out to be essential in improving current differentiation protocols for cell-replacement therapies.

2.4 ORIGIN OF THE PERIPHERAL NERVOUS SYSTEM

The peripheral nervous system is entirely derived from the neural crest. Neural crest cells are transient cells specific to vertebrates that constitute a versatile stem-cell pool capable of giving rise to numerous cell types and of contributing to different organs (Graham, 2003; Jessen and Mirsky, 2005). The cell types generated include autonomic and sensory neurons, Schwann cells (the myelinating cells of the peripheral nervous system) and chromaffin cells (the neuroendocrine cells of the adrenal medulla), but also cell types contributing to epidermal and connective tissues like melanocytes and cranial chondrocytes, osteocytes, adipocytes and dermal fibroblasts.

The formation of the neural crest is induced at neurulation and mediated by BMP signaling. Neural crest cells originate from cells at the border of the neural plate, which interact with the epidermal ectoderm that secretes BMP4 and BMP7. This interaction induces neural folding and, in turn, the neural crest. Molecularly, this induction is supported by the expression of a set of transcription factors including SLUG and FOXD3 (Graham, 2003).

After induction, neural crest cells transition from an epithelial to a mesenchymal phenotype: they free themselves from the dorsal neural tube, switching off the expression of N-CAM and N-cadherins, and become motile and delaminate (Theveneau and Mayor, 2012). After leaving the neural tube, neural crest cells migrate towards their sites of differentiation, following stereotypical migratory streams guided by environmental cues.

2.5 DEVELOPMENT OF THE SYMPATHETIC GANGLIA

Both the sympathetic and parasympathetic divisions of the autonomic nervous system are generated from neural crest progeny. The system is responsible for the maintenance of homeostatic condition, regulating body temperature, heart rate, respiratory rate, vasodilation, digestion and sexual arousal (Kandel et al., 2013). It acts on these physiological processes by releasing neurotransmitters to directly control smooth and cardiac muscle-fiber contraction and gland secretion. Most organs receive both sympathetic and parasympathetic input with the exception of sweat glands, adrenal glands, pilo- and nipple-erector muscles and blood vessels that receive only sympathetic innervation. Both components are modulated by inputs from the CNS that determine the activity of autonomic ganglia neurons.

At embryonic day 10, a group of ventrally migrating neural crest cells stops in the vicinity of the dorsal aorta. They aggregate to form a column of ganglion primordia, which extend rostrocaudally along both sides of the dorsal aorta and later coalesce to form a chain of sympathetic ganglia. During the process of migration and column formation, neural crest-derived cells become fate-restricted through the integration of extrinsic signals and intrinsic factors. The extrinsic factors responsible are bone morphogenetic proteins (BMPs) and WNTs (Hari et al., 2002; Schneider et al., 1999). BMP2, BMP4 and BMP7 are synthesized

and secreted by the dorsal aorta. These signals induce neural crest cells to differentiate into sympathetic neurons (Reissmann et al., 1996; Varley and Maxwell, 1996). In particular, BMPs play a gene-activation cascade whose members and mechanisms are well characterized. Two early transcription factors, ASCL1 and PHOX2B, top a regulatory hierarchy that includes the transcription factors PHOX2A, INSM1, HAND2 and GATA3 (Apostolova and Dechant, 2009; Goridis and Rohrer, 2002; Guillemot and Joyner, 1993). The coordinated expression of these transcription factors specifies neural progenitors to be noradrenergic sympathoblasts, which involves the activation of both pan-neuronal genes and cell-specific genes, such as the enzymes TH and DBH. Sympathetic neuroblasts begin to project axons and dendrites while en-route towards their destination. Axon projection occurs along the arterial vasculature towards target organs and is mediated by RET signaling (Kuntz 1934). RET signaling is activated by Artemin, which is secreted by the connective tissue and smooth musculature surrounding the ganglion to form a gradient. Artemin binds to its receptor GFRα3 and induces the formation of a complex that recruits RET and stimulates innervation. RET signaling is also essential for cell survival and subtype specification, although later this trophic support role is transferred to NGF by the upregulation of TRKA and the downregulation of RET (Birren et al., 1993).

"Therefore, either the reality on which our space is based must form a discrete manifold or else the reason for the metric relationships must be sought for, externally, in the binding forces acting upon it."

Bernhard Riemann - Lecture on the foundation of geometry – 1852

3 SINGLE-CELL RNA SEQUENCING

3.1 THE EVOLUTION OF THE TECHNIQUE

The preparation of cDNA libraries from single cells was first described by James Eberwine and Norman Iscove (Brady et al., 1990; Eberwine et al., 1992; Van Gelder et al., 1990). The foundational studies used either PCR or *in vitro* transcription to amplify the aproximatelly one picogram of mRNA contained in a single cell. Although cDNA analysis was limited to cloning and Southern blots, new important biological insight emerged. For example, the technique allowed Eberwine et al. to observe for the first time that morphology and electrophysiology do not correspond perfectly to the transcriptional profile of a cell and that molecular heterogeneity exists. These observations fostered further studies, and the technique gained some popularity, particularly in the neuroscience field, where the molecular heterogeneity of the tissue is especially high (Shumyatsky et al., 2002; Tanabe et al., 1998). Only later, with the availability of microarray technologies, was it possible to extend this approach and make it high throughput (Kurimoto et al., 2006). The first single-cell microarray was capable of detecting only a targeted fraction of known transcripts and was not able to discriminate splice isoforms or to obtain absolute quantification.

The first example of single-cell transcriptome sequencing was presented by the Surani laboratory (Tang et al., 2009). The focus of this pioneering work was to detect genes and splicing variants in the transcriptome of an individual cell, previously impossible with microarray technology. The analysis was gene-centric, an approach distinct from the cell-centric paradigm that characterizes the use of single-cell RNA-seq today. The important conceptual change was understanding the great value of single-cell RNA-seq for charting the high-dimensional landscape of gene expression. This realization motivated the development of the first method supporting multiplexing: STRT (single-cell tagged reverse transcription) (Islam et al., 2011).

STRT and subsequent methods introduced several improvements to the original technique from Tang et al.. These improvements included enhancing the efficiency of reverse transcription (RT), multiplexing the method to scale up the number of cells sequenced, and obtaining full-length coverage (Hashimshony et al., 2012; Picelli et al., 2013). A significant upgrade was the introduction of unique molecular identifiers (UMIs), barcodes incorporated into the cDNA during reverse transcription, that allow estimation of the absolute number of molecules present in each individual cell (Islam et al., 2013; Kivioja et al., 2011).

Reducing the cost of reagents and the requisite bench work was essential to improve throughput further. In this context, the next generation of single-cell protocols clearly had to scale up using microfluidics. One of the first solutions consisted of microwell arrays that could be loaded with both cells and barcoded beads (Fan et al., 2015). However, the initial

adoption of this technology was discouraged by the restricted platform and the lack of a detailed protocol to reproduce it.

Two landmark papers from Macosko and Klein popularized microfluidics-based approaches by introducing microdroplet-based single-cell RNA-seq (Klein et al., 2015; Macosko et al., 2015b). These techniques (Drop-seq and inDrop) encapsulate cells in monodispersed micro-emulsions and thereby significantly increase the number of cells processed. Each droplet contains a bead that is barcoded using a combinatorial split and pool strategy, which generates the millions of sequences required to reduce the chance that two identically barcoded beads are sampled. The adoption of these techniques brings the number of cells that can be processed in a couple of days to about 20-30k. Although these protocols are publicly available, the company 10x Genomics introduced the commercial "Chromium" platform (similar to InDrops) that contributed significantly to democratization of large-scale single-cell RNAseq. Initially, tinkering and microfluidics experience had been required to optimize the non-commercial solutions (Zheng et al., 2017).

More recently, a new set of methods using "*in situ* barcoding" has emerged. These methods scale single-cell techniques even further and tremendously reduce library preparation costs. They exploit the same combinatorial schemes used to barcode beads in droplet-based single-cell RNA-seq. However, instead of ligating the barcodes to beads, the split and pool steps are performed directly on the cDNA inside fixed and permeabilized cells (Cao et al., 2017; Rosenberg et al., 2018). Approximately 200k cells are processed in a couple of days, enough to obtain full single-cell atlases of small organisms.

3.2 ANALYSIS OF SINGLE-CELL DATA

In parallel to the technological advances described above, the scientific community developed several computational tools able to deal with the peculiarities of the new data. The challenges of analyzing scRNA-seq data are fundamentally different from those encountered analyzing classical gene-expression data. They are related to the structure of the data matrix, the interpretation of its entries, and the general aims of the statistical inference (Wagner et al., 2016).

To appropriately interpret the data matrix of a single-cell RNA sequencing experiment one has to consider the entities quantified and the process of sampling that generates the matrix. The experimental procedure samples mRNA molecules from the 10,000 - 500,000 molecules that a cell contains typically detecting 1000 - 20,000 UMIs. This number of UMIs is distributed over more than 20,000 genes. Furthermore, the cumulative influence of gene-specific levels of expression (e.g., highly expressed vs. lowly expressed genes) and technical bias (e.g., sequence-specific reverse transcription of PCR efficiencies) results in average abundances that can vary over four orders of magnitude (Islam et al., 2013). A single-cell

RNA-seq data matrix is therefore sparse, discrete and not trivially normalizable (Vallejos et al., 2017). This matrix will have to be carefully preprocessed and transformed to be useful to calculate the similarities (or distances) between cells.

Generally, single-cell analyses are a set of inference procedures performed on partial observations of a cell state (the columns of the matrix), with the aim of answering questions regarding the cellular and molecular composition of tissues. Note that this aim is fundamentally different from that of a classical bulk transcriptomics analysis, where efforts are directed towards controlling biases and noise to extract significant differences between samples (Vallejos et al., 2016).

Since the dispersion of single-cell RNA-seq data was analyzed for the first time, the raw data has been considered well modeled by a Negative Binomial distribution (Grün et al., 2014). However, analyses of homogeneous populations and depth-normalized data have shown that overdispersion is negligible and these data can also be modeled with a Poisson distribution (Islam et al., 2013; Ziegenhain et al., 2017). However, samples from these distributions will vary widely in their dispersions and this heteroscedasticity requires particular attention when performing feature selection.

Feature selection is a commonly used preprocessing step that consists in selecting genes that have a coefficient of variation higher than that expected by a background model. The expected coefficient of variation for each gene is estimated fitting a Poisson or a non-parametric model, using the mean as a predictor. Finally, genes with extreme residuals are selected, as they are likely to contain biological variation.

Several techniques have been developed to control for technical variation in single-cell RNA-seq data, especially with respect to factors that were not relevant to bulk RNA-seq. These factors include overamplification, zero inflation, cell doublets and variation in cell characteristics such as size and lysis efficiency (Finak et al., 2015; Kharchenko et al., 2014; L. Lun et al., 2016; Risso et al., 2014; Treutlein et al., 2014; Vallejos et al., 2015). Many approaches aim at extracting variance underpinning biological difference between cell types or states and excluding noise intrinsic to the process of transcription (Wagner et al., 2016).

An important preprocessing operation is the normalization of sampling depth and cell size. This is usually achieved by multiplying the counts of each cell against a scaling factor (Li et al., 2017; Vallejos et al., 2017). The procedure is necessary to bring all the samples to an appropriate scale for comparison. A difference of sampling depth might otherwise bias comparison between molecular profiles (i.e. using Euclidean distances). Different strategies can be used to estimate a scaling factor, some of which were devised for bulk sequencing data. However, methods designed for bulk data are less robust because of the characteristic zero inflation and sparsity of single-cell data. Some methods are biased because they tend to

rely on a few highly expressed genes, while others produce an overinflated scaling factor for cells rich with zero counts (Vallejos et al., 2017). A method crafted for single-cell data, which seems to significantly outperform earlier methods, deconvolves pool-based size factors into single-cell factors (L. Lun et al., 2016). A problem common to all normalization methods is determining an artifactual zero inflation, because the expected zeros of the Poisson distribution (false negatives) cannot be rescaled. A more drastic alternative is to skip depth-scaling altogether. This approach is possible but requires the use of a distance, such as correlation distance, that is not influenced by linear scaling.

Because raw single-cell data is distributed Poisson (or negative binomial), the variance associated with each gene scales with the gene's average expression level. This relation represents a problem when comparing gene expression in a Euclidean space. The noise of highly expressed genes causes a greater displacement in this space than lowly expressed genes, complicating the analysis. This problem is usually attenuated by using a variance-stabilizing transformation: a non-linear function that equalizes the variance across different level of expression (Anders and Huber, 2010; Love et al., 2014; Marioni et al., 2008). The most commonly used transformation is the logarithm function; however, alternatives have been proposed that avoid variance inflation at low expression levels (Wagner et al., 2018).

Another aspect of preprocessing aims to reduce batch effects and systematic technical errors, shielding downstream analysis from any variance that is not biologically relevant. A naïve procedure to address this problem is to "regress-out" the bias by fitting a linear model using batches or other factors as predictors, then using the residuals for analysis (Johnson et al., 2007; Ritchie et al., 2015; Satija et al., 2015). However, in single-cell RNA-seq datasets, the non-uniform cell composition between replicates can result in the generation of artifacts. To address this problem, methods have been developed that can merge datasets from different conditions, technologies and species (Butler et al., 2018; Haghverdi et al., 2018).

Discreteness and sparsity can obscure gene-gene correlations and render downstream algorithms less effective, especially if these algorithms expect continuous and normally distributed data. A solution is aggregating data over either cells or genes. Nonparametric methods, which aggregate read counts over small subclusters and then use the aggregated "metacells" for analyses, have been used to tackle particularly challenging datasets (Sebé-Pedrós et al., 2018a, 2018b). Aggregation on genes is instead aimed at generating variables (considered "metagenes") that vary smoothly and have less skewed distributions. These methods include principal component analysis, non-negative matrix factorization or more sophisticated latent variable models (Buettner et al., 2017; Lee and Seung, 1999). Finally, other methods preserve single-cell and single-gene distinction by locally sharing expression information among similar cells, and, therefore, behave analogously to a smoothing filter. One way to achieve this smoothing is averaging the expression of kNN. More complex

alternatives include simulating a diffusion process on the kNN graph or predicting with a statistical model fit on the data (van Dijk et al., 2018; Huang et al., 2018; Li and Li, 2018; Wagner et al., 2018).

The relation between the quantity of RNA and its functional role in the cell is not trivial. First of all, conversion to cDNA and amplification exhibits sequence-dependent bias (Islam et al., 2014). Furthermore, evaluating the correlation between the abundance of a particular mRNA species and the amount of translated protein is challenging. Even more challenging is determining the functional impact of a certain concentration of RNA. For example, even if a transcript is detected as low as one molecule in every ten cells, this low expression, over time, is often sufficient to sustain the translation of a functional protein.

Finally it is worth mentioning that, while the computational toolbox has expanded over the years, the advent of large-scale single-cell RNA-seq has impaired the adoptability of some methods (Klein et al., 2015; Macosko et al., 2015a; Zheng et al., 2017). The scale-up of scRNA-seq to tens of thousands of cells bears significant advantages but introduces a non-trivial computational burden. Every algorithmic procedure that scales polynomially, or worse, with the number of samples becomes prohibitively slow when applied to datasets of hundreds of thousands of cells. To comply with computational demands, many algorithms are already being replaced by more computationally scalable approaches. For example, new methods avoid calculating full distance matrixes and use approximate nearest neighbor graphs (Iacono et al., 2018; Wang et al., 2016; Wolf et al., 2018).

3.3 DATA ANALYSIS FOR CELL TYPE DISCOVERY

After normalization and feature selection, the first step of many single-cell RNA-seq pipelines is to group the data into subpopulations. This procedure is instrumental for further analysis, because it provides relatively homogeneous groups to work with. Grouping single-cell expression profiles into categories is a problem of clustering in a high dimensional space (Banerjee and Chaudhury, 2010).

Clustering is an ill-posed problem whose optimum is application-dependent. Ideally, for cell type discovery, we would like to group cells on the basis of the presence of regulators and their interactions with regulatory elements that promote a steady state. To adopt this strategy, we would have to know the concentrations of the relevant transcription factors, their affinities for different regulatory elements, and the availability of the latter. Currently, these systematic biochemical measurements are not possible, and clustering must rely on gene-expression profiles alone. Interpreting the relevance of different displacements in expression space would be enough for a principled clustering (i.e. How far is state B to A? What is the shortest path to reach state C?). But this idea fundamentally reduces to the same problem, requiring knowledge of the constraints imposed by internal regulatory machinery.

Due to our current lack of knowledge, a more naïve approach is commonly used: genes are considered the same as any other feature, and general-purpose machine learning algorithms for clustering are applied. An exhaustive comparison of clustering algorithms and their performance on diverse datasets is challenging. In the absence of absolute benchmarks, numerous clustering algorithms have been developed and claimed to best suit the distribution of scRNA-seq data and/or the peculiarities of specific datasets. However, many of the clustering algorithms crafted for single-cell data can be considered adaptations of well-known algorithms, rather than radically new approaches.

Many clustering methods are based on classical algorithms such as k-means or hierarchical clustering. An example is *RaceID* that combines k-means with outlier detection, which is effective for finding rare populations but performs poorly on classical datasets (Grün et al., 2015; Lin et al., 2017). K-means and other clustering methods tend to produce clusters of homogeneous size; approaches like *pcaReduce,* based on hierarchical clustering, instead yield clusters of different sizes (žurauskiene and Yau, 2016). Other methods like *ACCENSE* include a dimensionality reduction step, performed by PCA or t-SNE, followed by a density-based clustering approach. This method avoids the "curse of dimensionality" at the expense of some biological variation. Others, like BackSPIN, refrain from dimensional reduction and implement an iterative biclustering procedure (Gokce et al., 2016; Tasic et al., 2016; Zeisel et al., 2015).

The introduction of microfluidic technology encouraged methods that scaled better with the number of cells. A naïve speedup can be achieved by restricting the clustering problem to a low-dimensional space. For example, some of the first papers with tens of thousands of cells performed density-based clustering such as DBSCAN in a low dimensional space calculated with non-linear embedding methods such as t-distributed stochastic neighbor embedding (t-SNE) (Ester et al., 1996; der Maaten and Hinton, 2008). However, t-SNE was originally designed as a visualization approach, and its use in clustering pipelines is debatable. Dimensionality reduction techniques that provide more guaranties on preserving the global structure of the dataset, such as UMAP (Uniform Manifold Approximation and Projection), should be preferred (McInnes and Healy, 2018).

Graph-based methods such as the Louvain community-detection algorithm are very effective and extremely efficient with a large sample size (Blondel et al., 2008; Lancichinetti and Fortunato, 2009). They are also very versatile, effectively representing the non-linear structure of the manifold and identifying clusters of different sizes and densities. However, these clustering methods optimize a global function (modularity) on a graph. All such methods suffer from an inherent resolution limit, which precludes the detection of clusters smaller than some fraction of the graph's total size (Fortunato and Barthélemy, 2007). So-

called ”resolution limit-free” algorithms instead enable the detection of small clusters, but might unduly split large clusters into many fractions (Lancichinetti and Fortunato, 2011).

Despite the heuristic nature of the algorithms used, assessing the significance of the clusters obtained is challenging. The consensus of several clustering methods can be used to reinforce evidence of separation between cell types, but it sacrifices power; one algorithm might detect rare clusters missed by another. An alternative strategy involves sub-sampling and assessing the consensus among samples. This statistical procedure has the potential to eliminate the effects of low-frequency outliers or doublets. Furthermore, density-based resampling can reduce the bias introduced by non-uniform tissue composition and preprocessing procedures such as feature selection and clustering (Joost et al., 2016; Tasic et al., 2016).

After classifying cells into cell types, differential expression analyses or regression models are usually applied to determine cell type-specific expression patterns. Validation of the discovered cell types by orthogonal technology remains fundamental. To this purpose, *in situ* detection technology is preferred, and single-molecule fluorescent *in situ* hybridization (smFISH) has been the golden standard. These methods do not suffer from the problems associated with the use of antibodies, including non-specific binding and the scarce availability for some model organisms.

3.4 EXTENSION OF THE ANALYSIS TO DYNAMICAL SYSTEMS

Beyond tissue mapping, a more ambitious goal of single-cell analysis is the description of dynamic biological processes and the elucidation of gene-regulation mechanisms.

The challenge of studying these phenomena lies in the destructive nature of single-cell measurement, which provides only a static snapshot of cell states instead of a continuous time series. However, because cells in development are not synchronized, single-cell RNA-seq captures a continuum of different states that can be reordered to describe a typical cell development trajectory. This is the core idea of “pseudotime analysis” or “trajectory reconstruction”, an approach that has become standard for studying development at the single-cell level (Bendall et al., 2014; Trapnell et al., 2014). Since the publication in 2014 of Monocle and Wanderlust, the first two algorithms for trajectory reconstruction, 50 additional methods of trajectory inference have been released in only four years (Bendall et al., 2014; Saelens et al., 2018; Trapnell et al., 2014). The great interest that this approach has attracted is related to the importance of the biological problems it promises to solve.

The key steps of trajectory reconstruction are dimensionality reduction, clustering or graph construction, and the ordering of cells along a simplified representation. Monocle, for example, builds a minimum spanning tree in a low-dimensional embedding provided by independent component analysis (ICA). The cells are then ordered and the pseudotime between them considered proportional to their distance in ICA. Monocle uses the

pseudotemporal dimension to perform a time-series-like analysis (Figure 6). Wanderlust, another pioneering algorithm, makes use of a kNN graph to effectively counter the curse of dimensionality. The graph circumvents dimensionality reduction, because statistics only take into account the number of edges between cells. Subsequent methods like Wishbone extend trajectory analysis to incorporate branching processes, more appropriately modeling developmental datasets where a common progenitor stems into more than one fate (Figure 6) (Setty et al., 2016).

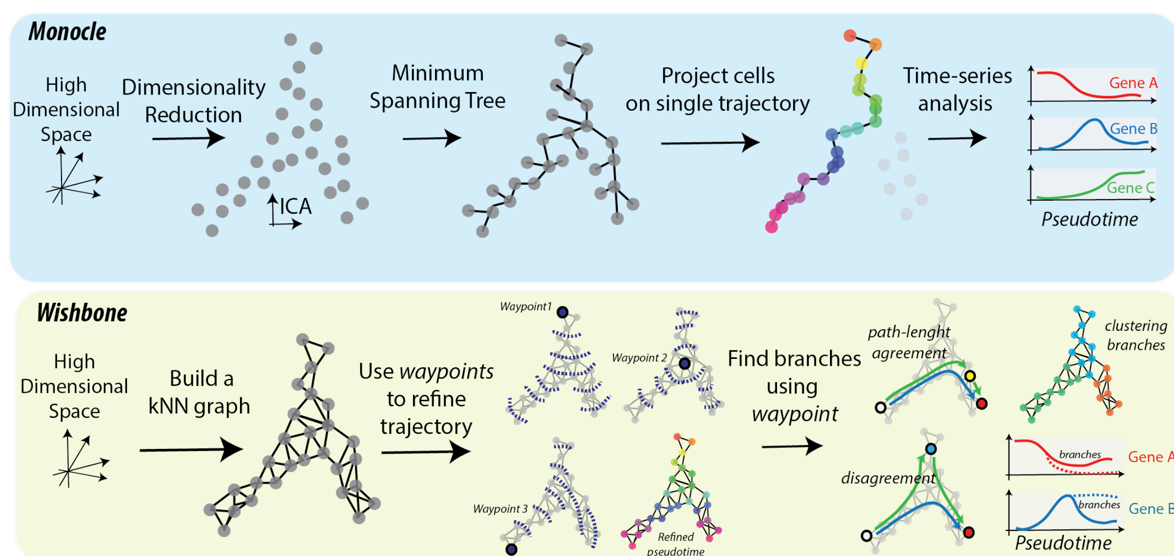


Figure 6 – Overview of two pioneering trajectory inference algorithms.

Different methods for pseudotime analysis use different models and algorithmic strategies and often require different prior information. For example, the user may need to specify the cell types to use as a starting point or the underlying topology to reconstruct. Furthermore, different methods have different limitations, such as their abilities to predict multiple branching or cycles (Saelens et al., 2018). Similar to clustering methods, trajectory-reconstruction methods must scale with recent increases in throughput and take advantage of the increased number of cells sequenced. Some methods use cell aggregation to computationally simplify the task while increasing robustness. A promising method, which uses this approach, is approximate graph abstraction (AGA) where cells with high similarity are grouped in a local pool. The flexibility and simplicity of the graph representation has many advantages; one is the ability to deal with disconnected graphs and therefore avoid forcing outlier groups into the trajectory (Wolf et al., 2017).

All trajectory-inference methods infer gene-expression dynamics from branching and non-directional gene-expression distributions. Strictly speaking, this inference is an underdetermined estimation problem: for any distribution observed, there are multiple dynamics that could explain it equally well. The work of Weinreb et al. delineates this indetermination, defining the minimal assumptions necessary to pose the problem well and

estimating the velocity of cells that traverse a manifold using the principles of mass conservation (Weinreb et al., 2018). Another recent method that tries to solve this problem is Waddington Optimal Transport, which uses advances in the mathematics of optimal transport to infer dynamics from data collected at multiple time points (Schiebinger et al., 2017).

Finally, it is important to note that these estimation methods rely on sufficient sampling of the transition between different states. If an intermediate exists for a short time, shallow sampling will detect an interruption of the continuum in expression space. Even when transition states are sampled sufficiently, sampling each stage to a different depth can bias the analysis results. For example, the pseudotime result might be strongly and non-linearly compressed or dilated with respect to real time. Furthermore, because of their strong priors, these methods will tend to find spurious transitions when applied blindly on relatively similar but dynamically unrelated cells. This becomes a greater concern when considering that embeddings of high-dimensional data can generate artifactual one-dimensional patterns. This is a well-known effect referred to as the "horseshoe phenomenon" and it is related to the inability to estimate long range distances (Diaconis et al., 2008; de Leeuw and Leeuw, 2007).

The error lies in seeking to understand the world in terms of things rather than events. It lies in ignoring change. The physics and astronomy that will work, from Ptolemy to Galileo, from Newton to Schrödinger, will be mathematical descriptions of precisely how things change, not of how they are. They will be about events, not things.

Carlo Rovelli - The Order of Time - 2018

4 THE CYBERNETICS OF DEVELOPMENT [*]

4.1 TIMESCALES AND RHYTHM

For embryonic development to successfully give rise to a complete organism, a significant number of events and interactions need to happen in a precise order. The speed and tempo of these events is fundamental for the correct shaping of tissues. Furthermore, timing is important to coordinate crosstalk between each tissue and the rest of the organism, as the rate of each morphogenetic process influences the others. Development requires remarkable coordination that needs to be achieved without rhythm-imposing physical phenomena (i.e., a zeitgeber) (Landgraf et al., 2014). If the embryo does not use an external clock, cell-autonomous mechanisms need to control developmental timing and constitute an internal molecular clock for the embryo.

There are several known examples of how clocks can be implemented in early embryonic development. One of the simplest is a countdown by dilution. In this mechanism a chemically stable inhibitor, initially present in the egg, is then diluted by segmentation and cell growth, allowing the inhibited process to start with a delay after fertilization (Collart et al., 2013).

Another common mechanism involves the timing of chained molecular processes. In particular, an auto-regulation mechanism combined with the half-lives of pre-RNA, mRNA and proteins effectively generates oscillations. An example is the delayed negative feedback circuit regulated by Hes and responsible for the somite-segmentation clock (Gibb et al., 2010). Although the segmentation clock is conserved in vertebrates, its frequency underwent adaptations. Segmentation times therefore vary more than one order of magnitude across vertebrates. An analysis of the molecular processes involved suggested that the different time scales can be explained by the kinetics of Hes splicing and nuclear export (Hoyle and Ish-Horowicz, 2013; Hubaud and Pourquié, 2014).

More generally, natural selection can act on different phenotypes by modifying the reaction rates and dissociation constants of the molecules involved in development. The change of each reaction rate can influence the size and shape of tissues, as well as set the speed of different processes (Ebisuya and Briscoe, 2018). Often, timescales need to be modified in a way that leaves the rest of the system invariant. For example, the homology of neuronal types in different mammals implies that genetic regulation networks are highly conserved.

[*] The chapter title is a tribute to a homonymous chapter in Waddington's book `*The Strategy of the genes*`. In the chapter he describes the development as a dynamical system and makes the famous analogy with a ball rolling on a landscape. The term *Cybernetics* is a, less popular, synonym of *Systems Biology*.

However, the timing of appearance and differentiation of these neurons can differ significantly between species (Van den Ameele et al., 2014; Davis-Dusenbery et al., 2014).

In primate evolution, differences in proliferation versus differentiation of cortical progenitors have been held responsible for the expansion of neocortex in humans. Recent studies comparing humans to chimpanzees suggest that this expansion might have been achieved mainly through a prometaphase-metaphase lengthening, since other molecular and cytoanatomical parameters do not vary (Mora-Bermúdez et al., 2016; Otani et al., 2016). It is remarkable and maybe surprising that such a simple change in timing, through the regulation of cell-cycle genes, has been critical for our cognitive abilities.

4.2 LINEAGES

The term *lineage* has multiple meanings in biology. In developmental biology, the term indicates both a mitotic kinship between cells and a progression of cells through a series of developmental states (Figure 7). In certain circumstances, these two meanings correlate, blurring this distinction and sometimes producing fallacious reasoning. In this section, I discuss how the concept of lineage is fundamental for understanding development. I highlight the differences between the two meanings of lineage and explain when they overlap. Finally, I will give an overview of both traditional and new technologies that can be used to study lineages.

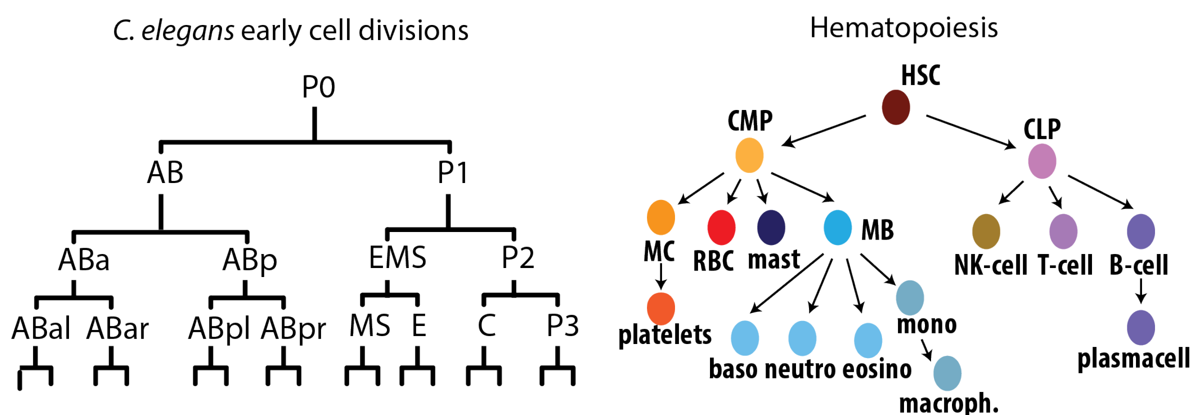


Figure 7 – Different meanings of *lineage*: cell divisions and progression of cell states.

4.2.1 Lineage as mitotic kinship

The genetic meaning of the term lineage relates to the idea that cells descend from the same mother cell through a certain number of mitotic divisions. Since divisions are binary by nature, this definition of lineage can be represented by a binary tree that describes the genealogy of cells.

In organisms consisting of a limited number of cells, like Nematodes, the total number of divisions that each adult cell undergoes is remarkably small (at most ten in *C. elegans*). For *C. elegans*, in particular, mitotic lineage relationships between cells are the same in different

individuals. Exploiting this fact, in the early '80s, the mitotic kinship between all cells in *C. elegans* was delineated by visually tracking cell divisions (Sulston et al., 1983). It is likely that the evolution of a stereotypical lineage was dictated by the necessity to build an anatomically complex and functional organism with a small number of cells.

For bigger animals, a greater number of cell divisions occurs, and a significantly higher number of cells is generated. As a result, chaos necessarily emerges from the large number of components and the presence of nonlinearities (Strogatz, 1994). Nevertheless, with an appropriate set of developmental mechanisms, the anatomy and histology of an organism remain invariant at the mesoscopic scale that is relevant for physiology. Under this premise, the fact that lineage is not stereotypical should not influence fitness, and the process is unlikely ever to be under selective pressure. Furthermore, the number of cells in an adult organism is so large that cells cannot be unequivocally identified. Therefore, for practical reasons, the description of lineage will be a statistical one.

The individual lineages are not relevant *per se*. Instead, a more satisfactory description can be achieved by asking: (1) What is the contribution of a particular group of cells A to a group of cells B? (2) How many other groups of cells contribute to a particular group of cells? These questions have traditionally been addressed using lineage-tracing technologies. Relating one molecular state to a later one on the basis of a mitotic relationship is such a standard procedure that it has probably contributed to the confusion of the two different meanings of "lineage".

4.2.2 Lineage tracing

Different approaches have been employed to study lineages. Overall they can be distinguished in *prospective* lineage tracing and *retrospective* lineage tracing.

Prospective lineage tracing methods were adopted earlier and remain more commonly used. They consist of labeling a single cell or a set of cells with a dye and observing the dye distribution in the embryo at later time points to identify the cells' progeny (Jacobson and Hirose, 1978). In this way, the first fate maps of presumptive tissue were constructed (Eagleson and Harris, 1990; Kimmel et al., 1990). For example, in *Xenopous laevis*, single blastomeres at the 32-cell-stage have been tracked and assigned particular fates (Moody, 1987a, 1987b). To adapt the strategy for organisms with a higher number of cell divisions, the dyes were substituted by transgenic reporters, which are integrated into the genome and inherited by daughter cells (Turner and Cepko, 1987).

Nowadays, the most commonly used approach relies on Cre-loxP genetic recombination. The technique was developed in the '90s to study lineages in Drosophila. It uses a recombinase, often placed under the control of a cell-type specific promoter, to activate the expression of a conditional reporter gene (Harrison and Perrimon, 1993). Recombination is achieved by

crossing a Cre mouse line with a reporter line that has a flanking loxP-STOP-loxP site. (Kretzschmar and Watt, 2012) Furthermore, the timing of the recombination can be controlled using Cre recombinase fused to the estrogen receptor (Feil et al., 1997).

Retrospective lineage tracing takes advantage of naturally occurring mutations in cells to reconstruct lineages. This approach advantageously does not require any intervention or preparation and can therefore be used in wild-type animals and human samples. However, these techniques usually require high sequencing depths, since rates of somatic mutation are low (Milholland et al., 2017). The problem can be attenuated by focusing on copy number variations (CNV), which can be obtained with shallow sequencing (Cai et al., 2014). In particular, CNVs accumulate quickly during tumor progression, making this kind of tracing particularly useful to study cancer (Wang et al., 2014). Apart from the demanding sequencing requirements, another limit is the difficulty of inferring intermediate clonal composition, since only leaf cells are analyzed, and time information is absent from the data.

4.2.3 Lineage as the progression through molecular states

The term lineage can reference a succession of molecular states that cells traverse to reach a mature state during differentiation. This concept of “differentiation lineage” can also be represented as a tree, similarly to what is done for mitotic lineage, or more flexibly with a directed graph. However, if a tree representation is used, the tree is not a binary one and has little in common conceptually with the mitotic genealogy. Cells at the root of the tree represent early states that are able to progress (depending on stimuli or stochastically) into more mature, less potent states towards the leaves (Figure 7, right). Therefore, it is implied that cells cannot move up the tree but only progress in one direction irreversibly; the root represents totipotent cells, such as the fertilized egg and early blastomeres, and the leaves are terminally differentiated cell types. This representation is summarizes differentiation as a series of states where each state progress into the next (Banerji et al., 2013). In contrast to the mitotic lineage, this progression is invariant within a species: at least to a certain level of approximation, it is the same across all individuals. Variation between individuals (genotype differences) and single cells (intrinsic noise) exists at fine granularity but only marginally affects the progression.

When we talk of lineages as a progression through states, we refer to a process controlled by regulatory machinery that can be only very approximated described by a tree, graph or any step-wise schematics. Ultimately this lineage is described by a multitude of biochemical rates and interaction laws that characterize the cell as a physical, dynamical system (Alon, 2007). A mathematical model of the process is introduced in the last paragraph of this section.

In practice, the two meanings of lineage are at least partially correlated for different reasons. First of all, symmetric cytokinesis partitions cell contents equally between daughter cells so

that the daughter cells share the same molecular state. Secondly, because daughter cells often remain in close proximity, they will respond similarly to morphogenetic gradients and are more likely to interact with the same neighboring cells. However, cell migration within the tissue and the formation of sharp boundaries by morphogens can easily reduce this correlation. Therefore, contrary to what is sometimes stated, the mitotic lineage is not fully embedded in the differentiation lineage (Marioni and Arendt, 2017). Such an embedding is consistent only if we think of lineage progression as fully autonomous and deterministic. Using somatic mutations accumulated during development to perform a clonal analysis, one study showed that adult neurons are often more closely related to a heart cell than to any of their neighboring neurons (Lodato et al., 2015).

4.2.4 Combined lineage-phenotype analysis

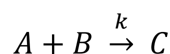
Recently a series of techniques have emerged that combine lineage (or clonal) tracing with high-throughput phenotyping schemes, including single-cell RNA sequencing and multiplexed smFISH (Frieda et al., 2017; Kester and van Oudenaarden, 2018). The data generated by these techniques promises a systematic way to combine lineage-tracing information and phenotypic data.

One method uses the CRISPR system to generate a combinatory diversity of mutations that accumulate during development. This is achieved by designing a guide RNA directed against an array of target elements inserted into the genome. I this way, each target element can be cut, repaired by non-homologous end joining and produce inheritable mutations (Mckenna et al., 2016). In three recent papers, this CRISPR-based lineage-tracing approach has been combined with the acquisition of single-cell transcriptomics data (Alemany et al., 2018; Raj et al., 2018; Shapiro, 2018; Spanjaard et al., 2018). These first implementations of a lineage recording system were applied to the model organism *Danio rerio,* the zebrafish. The methods necessarily take advantage of the animal's external fertilization, copious brood and large embryos to guarantee a reliable injection of the CRISPR-Cas9 components. The only example of this approach in the mouse was presented by the Church lab. The group generated a mouse line that uses a self-targeting CRISPR-Cas9 system, where the guide RNA can introduce mutations in its own sequence (Kalhor et al., 2017, 2018; Perli et al., 2016). These techniques appear promising as they make it possible to analyze cellular transcriptomes knowing, at the same time, the individual lineage tree that generated them. However, a recent simulation-based power analysis suggested that all of these techniques yield trees of very low accuracy and resolution, suggesting that the data from this techniques should be interpreted conservatively (Salvador-Martínez et al., 2018). These limits seem to be recognized by other authors who suggest that the technique provides coarse-grained clonal information, which can be refined using trajectory-inference methods (Kester and van Oudenaarden, 2018).

Interestingly CRISPR systems are not used only in lineage tracing. Recently the approach has been extended to record signaling events, particularly Wnt-dependent expression of a guide RNA (Frieda et al., 2017). I believe that this kind of signal recording is at least as interesting as lineage information, enabling studies of how signaling and cell-cell interactions affect single-cell transcriptomes.

4.3 MATHEMATICAL MODELING OF DEVELOPMENT

The molecular state of a cell is defined by its chemical composition. The concentrations of all chemical species in a cell can be considered a point in multidimensional space, where each axis corresponds to the concentration of a different molecule. Enumerating the vast variety of molecules that constitute a cell including nucleic acids, proteins, and metabolites is an near-impossible task, especially considering the existence of different variants for the same molecules (conformations, posttranscriptional modifications). However, we can make a very rough approximation and focus on the species involved in gene regulation. In particular, we can consider the transcripts and proteins that regulate transcription (transcription factors and other proteins that can interact with them). Modeling the bimolecular interactions between these molecules is a central aim of systems biology (Alon, 2007; Stumpf et al., 2011). Usually, in systems biology, the set of interactions between molecules is described with a system of ordinary differential equations that describes the temporal evolution of cell composition (Chen et al., 1999; Gardner et al., 2015). The differential equations are obtained by considering the reaction mechanism and the law of mass action. For example, let us consider the following reaction:



$$A + B \overset{k}{\rightarrow} C$$

If we assume that the system is well mixed, we can write the rate of change of the chemical species C as a function of the concentrations of the reagents that produce it:

$$\frac{d[C]}{dt} = k\,[A][B]$$

Chemical reactions in biological systems are for the most part catalyzed, the interactions between proteins can be cooperative and the available binding sites can be often saturated. For these reasons, correlations between concentrations and rates should be represented by non-linear "input functions"; a commonly used one being the Hill function (Alon, 2007). For instance, a simple gene regulatory network involving mRNA and protein species can be represented using the following system of ordinary differential equations:

$$\frac{dr_i}{dt} = \alpha_i f_i(\boldsymbol{p}) - \gamma_i r_i$$

$$\frac{dp_i}{dt} = \sigma_i r_i - \lambda_i p_i$$

where we enumerate with the index i the different genes and their respective abundances and parameters (α is the maximum transcription rate, σ is the translation rate, and γ and λ are the degradation rates). f_i indicates a generic input function (e.g. the Hill function), representing the effect of each transcription factor's concentration on the transcription rate of a gene. A common approximation for small perturbations around steady states is considering f_i linear. By rewriting the system of equations in matrix notation we obtain the following linear dynamical system:

$$\dot{\boldsymbol{x}} = A\boldsymbol{x}$$

Most efforts in systems biology are aimed at reverse engineering the gene regulatory network by developing ways to estimate the parameters of this equation (e.g. the matrix A or complex functions operating on the vector $\boldsymbol{x}$). Knowledge of these parameters would allow the simulation of a new situation (e.g. from any given starting condition).

This equation describes the evolution of the system (i.e. transcriptional activation and repression events) as a movement in high-dimensional space. Therefore, the observed trajectories in single-cell RNA-seq datasets are closely related to the solution of this equation. Furthermore, this model makes explicit that the time derivative of gene expression, relevant to the results of this thesis, can be interpreted as the results of a regulation-related operator acting on the current state of the cell. This insight is important to introduce paper 3, where we developed a method to estimate the derivative of gene expression $\dot{\boldsymbol{r}}$ at a point $\boldsymbol{r}$ in expression space.

It is interesting to notice the close relation between this model and a popular analogy created by Conrad Waddington in the '50s that compares a differentiating cell to a marble rolling in an undulating landscape (Figure 8, right) (Waddington, 1957). Waddington's analogy stemmed from the intuition that, in development, not all the structures are preformed in the embryo (preformation theory) and just need to grow in size, but interactions between the constituents of the embryo were responsible for the adult phenotype (epigenesis).

Despite that little was known about the biochemistry of gene expression, Waddington realized, ahead of his time, that the appropriate way to model this interaction was through differential equations describing a movement in an high dimensional phase space (Figure 8, left). He drew the famous epigenetic landscape as a simplified representation when he realized that: "A multidimensional phase space is not very easy […] to imagine or to think about" (Deans and Maggert, 2015; Waddington, 1957).

It is worth noting that the term "epigenetic" was only later co-opted to denote "nuclear inheritance, which is not based on differences in DNA sequence" and was instead defined by Waddington as "interactions between genes and their products which bring the

phenotype into being." This clarification is fundamental to avoid confusion about what the landscape represents (Holliday, 1994; Wu and Morris, 2001).

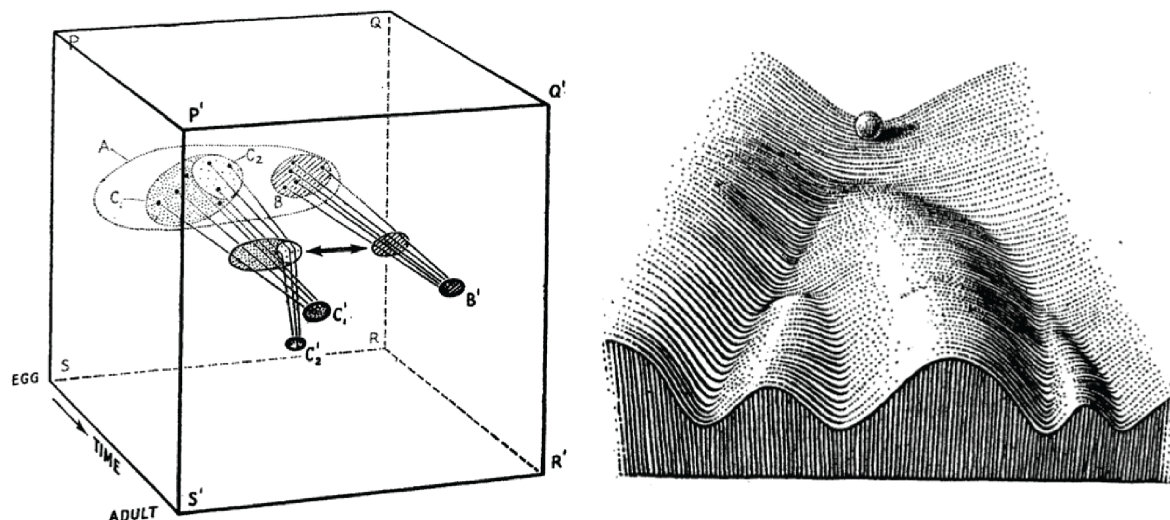


Figure 8 – Waddington's phase space diagram (left) and, its simplification, the potential landascape (right).

The system of differential equation described above is an unrealistic model: the system is closed and does not take into consideration any external influence. The model is only a good approximation in a short time interval where we can assume no external intervention is taking place. It is, therefore, only reasonable in situations where we can think of time evolution as fully autonomous. In a more realistic situation, external stimuli, signaling between cells and variations in the microenvironment can influence the time evolution of a cellular system. The model needs then to be extended to include both external stimuli and intracellular expression:

$$\dot{\boldsymbol{x}} = A\boldsymbol{x} + B\boldsymbol{u}$$

where $B\boldsymbol{u}$, called *input term* in control theory, describes which inputs $\boldsymbol{u}$ can be used to perturb the state of a cell and B describes how tuning these controls affects the rate of change of each molecule.

External interactions are highly relevant to shape the differentiation trajectory and therefore should not be disregarded when formulating a model. Furthermore, to describe the evolution of any cell accurately, one has to take into account other cells and the tissue as a whole. This extension requires incorporating positional information into the model and dealing with the diffusion of morphogens. The extension can be rewritten by adding a partial derivative with respect to the spatial coordinates:

$$\frac{\partial x_i}{\partial t} = f_i(\boldsymbol{x}) + g_i(\boldsymbol{u}) + D\nabla^2\boldsymbol{x}$$

Such a model would be even more appropriate if we considered compartmentalization and described the cells as a lattice (Olimpio et al., 2017). A broadened model could then include the effect of epigenetic modifications that stabilize such regulatory networks by complicating the transcriptional reactivation of methylated DNA (Bintu et al., 2016).

Thinking about a cell as a high-dimensional dynamical system is useful for reflecting one final time on the idea of cell types. It is tempting, in this sense, to assimilate the idea of a cell type to the one of an "attractor". The existence of attractors is one of the many achievements of dynamical-systems and chaos theory (Strogatz, 1994). The idea of a cell type as an attractor was proposed and elaborated by Stuart Kauffman in the context of boolean network modeling of gene regulation (Kauffman, 1969). We could think about a cell type as a fixed-point attractor of the non-autonomous evolution of the molecular system; a point approaching close enough is guaranteed to fall into it in a stationary or oscillating state. The idea of attractor can also accommodate more complex situations, such as cases where gene expression continues oscillating but remains constrained inside a well-defined region of expression space. Interestingly this notion nicely reconciles the intrinsic chaotic nature of a high-dimensional non-linear dynamical system with the fact that development appears stereotypical despite the fact that identical initial conditions cannot be guaranteed.

Finally, it is essential to mention the role of stochasticity in these dynamics. The stochasticity originates from the intrinsic randomness of transcriptional processes like transcription factor binding and translational bursts (Yu et al., 2006). The importance of stochasticity has been extensively studied by a whole community of system biologists and physicist (Elowitz et al., 2002; Hilfinger et al., 2016; Raj and van Oudenaarden, 2008; Yu et al., 2006). For example it is known that the stochasticity in each regulatory step is often dampened by the time-averaging mechanisms of nuclear retention, high levels of expression, or network motifs such as a negative feedback loop. However, noise can also propagate and be amplified through the gene regulatory network, and more general mechanisms to avoid these fluctuations might be required (Li et al., 2018). Modeling a system with intrinsic noise requires a probabilistic description of reaction kinetics: the chemical master equation. Usually, these systems are approximated through Fokker-Plank or Langevin equations and appropriately simulated with Monte Carlo methods (Gillespie, 1977; Lei, 2011). In this thesis we do not venture into this more advanced modeling but recognize it as an interesting perspective for future studies.

4.4 RELATION BETWEEN DYNAMIC AND EVOLUTIONARY PERSPECTIVE

In this chapter, I presented a dynamical view of gene regulation and showed that cell types can be viewed as attractors. This dynamic description does not contradict the evolutionary definition presented earlier; rather, the two perspectives complement each other.

Thinking in terms of a dynamical model renders the evolutionary framework more quantitative and clarifies the parameters of gene regulation that are tuned by evolution to generate new cell types. Mutations affecting the strength of protein-protein and protein-DNA interactions alter the transfer function that maps the molecular state vector $\boldsymbol{x}$ to the time derivative $\dot{\boldsymbol{x}}$. In this way, changes in the biochemical affinities of molecules affect multiple aspects of developmental dynamics and the lineage progression.

Were the transfer function known exactly, dynamics could be fully predicted and steady states (cell types) of the system determined by analysis or simulation. In this ideal situation, an evolutionary analysis would only provide extra information on the historical aspects of how cell types arise.

In practice, parameters for all possible molecular interactions and therefore the transfer function may be impossible to determine. Phylogenetic analysis, revealing conserved and novel regulatory modules, offers an opportunity to effectively simplify the model. For example, once the important regulators are identified from their high conservation, many other less important interactions could be neglected.

5 RESULTS

5.1 PAPER I - VISCERAL MOTOR NEURON DIVERSITY DELINEATES A CELLULAR BASIS FOR NIPPLE-AND PILO-ERECTION MUSCLE CONTROL

In this work, we applied one of our earliest single-cell RNA-seq technologies to study the stellate and thoracic sympathetic ganglia at postnatal day 27-34.

Previous studies had failed to identify the full cellular heterogeneity of sympathetic ganglia, and they had described only two cell types: cholinergic neurons (marked by Chat) and noradrenergic neurons (marked by Th). In this work, clustering the single-cell data revealed seven molecularly distinct cell types: two cholinergic and five noradrenergic types.

Cholinergic cells expressed Vacht, Vip, and Chat but were distinguishable by the expression of other markers such as the Cck-b receptor and somatostatin. Noradrenergic neurons (named NA1-5) were uniquely defined by a combination of makers including Enc1, Gfra2, Gfra3, Rarres, and Ret. These markers were used to validate the discovery *in vivo* and to link the molecular pattern to function.

By retrograde tracing, we found that two classes of noradrenergic neurons, NA2 and NA5, innervate the nipple erection muscle (NEM) and the piloerection muscle (PEM), respectively. Using reporter lines, we showed that these two classes of erector muscle neurons are born embryonically but differentiate only postnatally at a time point corresponding to the organogenesis of the target. These neurons project axons early and navigate to the target site before organogenesis, but they start innervating after the organ matures.

Our single-cell analysis showed that NA2 and NA5 neurons re-activate Ret expression, which is down-regulated embryonically, and express the coreceptors GFRA3 and GFRA2, respectively. Consistently, we showed that NEM and PEM express ARNT and NRTN, strongly suggesting the involvement of neurotrophic factors in driving differentiation and target innervation of these neurons. We confirmed this hypothesis by knocking-out Gfra and Ret and showing that these perturbations affected innervation of the erection muscle.

Overall, our results described unexpected heterogeneity of the sympathetic system and provided evidence on how different gene-expression programs ensure axon projection to well-defined target tissues.

5.2 PAPER II - MOLECULAR DIVERSITY OF MIDBRAIN DEVELOPMENT IN MOUSE, HUMAN, AND STEM CELLS

Using single-cell RNA-seq, we performed a heterogeneity analysis of ventral midbrain development in both mice and humans. For the mouse, we molecularly identified 26 transcriptionally distinct cell states across the time span of E11.5-E18.5. In the human, we identified 25 cell states sampling from week 6 to 11.

Taking advantage of the UMI-based absolute quantification of transcripts, we established a pattern of expression for each cell type and used it to compare cell types between the two species. We found *bona fide* homologous cell types, matching cell types on the basis of pairwise correlations over homologous genes. Using these homologous cell-type pairs, we looked more closely at the nature of this expression-pattern conservation and compared developmental timing between species, using the time of appearance of each cell type. We found striking conservation of developmental timing in line with scaling laws previously described. Differences between the tissues included less proliferative human progenitors, which we hypothesize is a strategy to control cell production over longer gestation times.

Our analysis of radial glial progenitors detected an uncharacterized heterogeneity; we defined at least four classes of radial glia-like cells and performed sequential smFISH to map their spatial and temporal dynamic arrangements. Our data provide evidence for the existence of different states of radial glia-like cells, which are spatially segregated and starting to acquire neuronal, astrocytic or oligodendrocytic character.

Previous work had identified five types of adult dopaminergic neurons. We investigated whether the heterogeneity of dopaminergic cell states arises embryonically. We found that adult dopaminergic subtypes emerge postnatally from two embryonic types, and we tracked this conversion using immunohistochemistry.

Finally, we provided an example of how this data can be used as a resource to answer questions with clinical implications. In particular, we addressed the lack of quantitative methods to evaluate stem-cell differentiation protocols. We devised an approach based on a probabilistic classifier that, after training on a human embryonic dataset, is capable of classifying single-cells transcriptomes. We showed how the approach can be used to evaluate the quality of stem cell-derived dopaminergic neurons engineered for Parkinson's disease cell-replacement therapies.

The method is capable of revealing the heterogeneity of the cell preparation and it facilitates the identification of genes that the protocol failed to induce. We believe that this approach can offer researchers a way to quantitatively assay the quality of *in vitro* preparations and effectively explore the data, providing insights on how differentiation protocols can be improved.

5.3 PAPER III – RNA VELOCITY OF SINGLE CELLS

In this paper, we proposed a way to overcome a fundamental limit of single-cell RNA-sequencing data for the analysis of dynamical biological processes, the absence of a temporal dimension. The novel method we presented in this paper estimates the first derivative of gene expression for each gene in a cell, a concept that we refer to with the name "RNA velocity".

The technique is based on the observation that the timescale of developmental, regenerative and reactive processes in both humans and other mammals matches the timescale of RNA metabolism. Main events in development happen in the timespan of few hours and the characteristic time of RNA metabolism (transcription, splicing and degradation) is also on the order of hours. In particular, we exploited the fact that both the abundance of spliced mRNA and newly synthesized, unspliced pre-mRNA can be estimated from single-cell RNA sequencing. The abundance of spliced mRNA is informative of the transcriptional events that happened in the recent past, and unspliced pre-mRNA of the expression level in the immediate future.

From this idea we developed an algorithm that estimates the rates of change in gene expression and predicts the future expression levels of each cell. The approach transforms a typical "snapshot" transcriptomic dataset into a dynamic representation of a process. In particular the method associates to every sampled point in expression space $\boldsymbol{x}$ a time derivative $\dot{\boldsymbol{x}}$ that is the result of gene regulatory networks and interactions of the environment (as described in Chapter 4).

Therefore, "RNA velocity" provides a new measurable object with relevant conceptual value: a vector field in expression space that can be used to describe dynamical changes, including complex situations such as cyclical trajectories or opposing flows. Therefore, the RNA velocity field, a conceptual extension of the Waddington's potential landscape, is ideal to model commitment, fate choice, cell cycle and the transcription kinetics.

Finally, we also provided evidence that RNA velocity can be estimated in human embryonic tissue and showed that spliced-unspliced traces can be used to study gene regulation strategies (overshooting, rates of change etc.) in human development. We expect that the approach will allow acquiring lineage-tracing-like data from human embryonic tissues.

6 CONCLUSION AND PERSPECTIVES

Single-cell RNA sequencing is revolutionizing the field of genomics, histology and cell biology. The work presented in this thesis spans a crucial period for the evolution of single-cell transcriptomics. During this period, the development of methods for single-cell studies has been extremely fast, such that chemistries and computational approaches developed half a decade ago are already considered obsolete. Our work is an attempt to exploit this technological progress to answer questions of developmental neuroscience. These efforts resulted in several contributions to both fields. To the field of developmental neuroscience my work contributed three discoveries of significant impact: the unexpected heterogeneity of the sympathetic ganglia, the existence of distinct classes of radial glia-like cells in the embryonic ventral midbrain and the discovery that dopaminergic neuron diversity completely unfolds only postnatally. My contribution to the field of single-cell analysis consisted of providing the first systematic comparison of cell types across organisms and between embryonic tissue and *in vitro* differentiated cells. Furthermore, we discovered that unspliced reads can be used to extract temporal information from single-cell data. We estimated the global time derivative of gene expression, a quantity at the center of dynamical theory in systems biology and that previously could not measured in single cells.

To put these contributions into perspective, we need to consider that single-cell transcriptomics is progressively being adopted as part of the ordinary toolkit available to biologists. The advantages of such a diffusion are several and go in the direction of a more reproducible and quantitative science. The scientific community can collect, in a cumulative and decentralized way, vast datasets across tissues and species. Producing and analyzing new data and organizing it in a useful resource are goals of the Human Cell Atlas, an international project, with dozens of partners, with the mission of creating a reference map of the all human cells (Regev et al., 2017). The availability of such an atlas is expected to have an important impact on the way we do biology. A change that might even be comparable to the way the Human Genome project affected modern biology.

Some of the possibilities that the Human Cell Atlas will open are foreshadowed, on a smaller scale, by some of the analyses presented in this thesis. For example, an *in vitro* preparation will be readily comparable to an *in vivo* counterpart, allowing the development of *in vitro* systems better modeling pathologies. More generally, quantitative comparison of molecular phenotypes across laboratories will be facilitated. In some cases the discovery of unexpected heterogeneity in a system or model of interest will open possibilities for re-interpretations of previous evidence.

The newly discovered heterogeneity in the developmental nervous system is a first step towards exploring these systems without neglecting their complexity. For both the

sympathetic ganglia and the ventral midbrain, many questions on how cellular heterogeneity is achieved remain still open. To achieve new insight into these mechanisms, the scientific community could start focusing attention on the correlation of expression patterns with mitotic lineages or with extrinsic signals that cells receive during development. Interestingly, both perspectives should be readily addressable with new technologies that record information about lineage or signaling events in the genome, using CRISPR-Cas9 system.

Ultimately, to understand the developmental process in its entirety and move towards inferring mechanisms of gene regulation, we need to combine transcriptomics with other genomics measurements. Over the last few years, several techniques have been scaled down to accept a limited amount of input material, rendering them single-cell compatible. In some cases, appropriate technical adaptations allow simultaneous detection of more than one kind of variable. This emerging multi-omics paradigm promises to offer an accurate representation of the state of a cell. The combination of this information on the cell molecular state with the rate of change of gene expression (e.g. by RNA velocity) will help elucidating gene regulation mechanisms and enhance our ability to predict responses to perturbations.

7 ACKNOWLEDGEMENTS

Here it comes the part of the thesis that usually attracts the broadest audience. This ended-up being longer than usual, but I am leaving Sweden and maybe a bit nostalgic.

First of all thank you **Sten**, for your generosity, fairness and commitment, learning from you made of my PhD a privileged time, I enjoyed every bit of it. Thank you for granting me a lot of trust from the first moment, for listening to my ideas when they were raw and naïve, giving them more credit than they deserved. Thank you for creating such a unique work environment that your lab is and for really caring for each of its members.

Thank you **Ernest** for your supervision, our endless discussions and the great collaborations, I really enjoyed working with you and I learned a whole lot! I am really thankful for the trust you gave me during this time. Thank you **Peter Kharchenko** for your fairness and helpfulness, for being a tireless collaborator and for showing me how to work at a pace without compromising quality. Thank you **Patrik** for the collaborations, your career advices and the esteem. Thank you **Jussi Taipale** for inspiring me with your scientific foresight. Thank you **Giorgio** for the mentoring and the scientific discussions.

Thank you **John Marioni**, for taking your time to be my opponent, it is an honor! To **Jonas Muhr**, **Ana Teixeira** and **Per Lötsted** thank you for your expertise on my examination board. Thank you **Vicente Pelechano** and **Eva Hedlund** for the counsel during my half-time. Thank you **Carsten Daub** for the esteem and giving me opportunity to teach what I like.

Thank you the PIs of MolNeuro: **Ernest**, **Goncalo, Jens, Patrik, Per** and **Ulrika** for your effort in making our division one of the most productive, healthy and fair work environments that I know. Thank you **Igor** for inspiring me with your enthusiasm and passion for biology. Thank you **Maria K** for giving me an opportunity when I was just a newbie and for introducing me to Sten. **Simon J** for the discussions and the nice collaborations.

Thank you **Amit** for being a teacher and companion of adventure during these years. Your contribution to the atmosphere in the lab and my development as a scientist has been invaluable. Thank you also for being a great baker and the ultimate example of productivity, probably unreachable to me, but at least I know what to aspire to. Thank you **Alessandro F** for sharing obsessions and for your friendship. We started with you complaining for my intrusion in the office, continued with the "nipple collaboration" and we ended up listening to each other complaints, thank you for all of this. Thank you **Daniel** for the wonderful collaboration, tolerating me effortlessly in hectic times and for balancing my character with your tranquility. Thank you **Ruslan** for all the help, for the days in Boston and a couch to crush on.

Thank you to shared the office with me. Thank you **Elisa** for your stoic tolerance and admirable professionalism, to **Eneritz** for discussions on bioinformatics and to be last in the ranking where Karol is first. Thank you **Nathan** for the fervid discussions, to **Manuel** for bringing always a smile in the office, even in November. Thank you **Lauri** for listening to me, for sharing with me your interesting point of view for your advices on life, science and movies. **David** for reviving my passion for physics and being a great companion of procrastination, our ability to keep up a long chat is epic, I enjoyed it a lot.

Now the labmates! Let's start with the alumni. Thank you **Hannah** for your kindness, all the help and support and for making Amit wear long trousers, at least some times... **Una** for being an example of strong will and endurance. **Kasper** for sharing the fun of making PhD

movies. **Saiful** for prophetic words and for developing STRT, without it I don't know what this thesis would be about. **Job** for the discussions, the use of social media and the laughs. Now the current members! Thank you **Lars**, for your friendliness, kindness and party spirit, and for all the nice discussions at lunch; **Simone C** for all the advices, all the help on the bench, for being a great talent-scout and trainer and for rising the quality bar. **Emelie** for your help, your kindness and for listening to my babble (without believing it). Thank you **Kimberly** for the availability, the patience and the collaborative spirit. **Alejandro** for bringing Spanish spirit to the group, the hospitality and the best paella Valenciana. **Camiel** for sharing the appreciation for math and algorithms. **Miri** for giving a chance to this cold country, it is great pleasure that you joined the lab. **Lijuan**, for your precious help and keeping my gibberish under control. Thank you **Peter L** for the availability, the kindness, the music, the help and all the great work that is so fundamental for what we do. **Anna Jo** for your resourcefulness and for the efforts in keeping the lab happy, you are, in a sense, the most Sicilian Swede I met... I hope you take it as a complement.

Thank you **Anna Ju**, for a contagious a sense of tranquility; **Marcela**, for our cheerful chats, for all the help and for caring. Thank to you **Karolina**, **Simone P**, **Anastasius**, **Henrik** and **Samaneh** for keeping up the single-cell spirit.

Thank you **Dimitry** for your wonderful stories, speeches and toasts. Thank you **Enrique** for the laughs and nerdy conversations. Thank you **Boris** for your fantastic interpretations, your authenticity, wit and cheerfulness. Thank you **Karol** for making me only second in the ranking of the most penetrating voice in MolNeuro. **Kaneyasu** for the hard work and help, to **Yizhu** for the loughs, the trust and your modesty, to **Carmen** for your kindness and care. Thank you **Johnny**, **Jana**, **Tony** and **Ahmad**, for contributing in making MolNeuro a pleasant place to work in. Thank you **Alessandra** for you infinite patience, kindness and for helping me seeing the light at the end of the bureaucratic tunnel. **Mandy** for your cheerfulness and for taking care of Daniel. **Anna F** for your enthusiasm and warmth. **Alessandro B** for taking Karol pole position after he left. **Dagmara** for friendliness and occasional sarcasm. Thank you **Shigeaki** for sharing the passion for good food, feeding me Ramen and for the increase of sushi quality that is ensured by your presence. **Ivar** for the jokes and for teaching me that there are not forbidden ingredient mixes in a meal. **Shanzeng** for you inquisitive questions and your smile of agreement. Thank you **Hermany**, **Carolina** for your friendliness and party spirit and to **Carmany**… weird that in all these years and somehow we never crossed! Thank you all fantastic people of MolNeuro: **Josè**, **Sam**, **Marek**, **Martin**, **Alcha**, **Victoria**, **Fatima**, **Komgrit, Pia, Hind, Pawel, Spyros, Songbai, Yiwen, Carlos**, **Rashid**, **Milda**, **Mukund**, **Leipa**, **Ibrahim**, **Kasra**, **Anna Mu**, **Alexandra**, **Lili**, **Laura**, **Puneet**, **Dongoh**, **Jussi K**, **Daohua**, **Elsemieke**, **Michael**, **Mindong**, **Jie**, **Kai Wai**, **Emilia**, **Goran**, **Connla** and **Anneke.**

Now friends from the world outside the lab, some of which are now spread around the world. First and foremost thank you **Denis** and **Javier S** for giving me the possibility to live in an environment where I could focus fully on intellectual and personal growth. Thank you **Chema** for exposing my "lazy italian" attitude to as many people as possible, your friendship and advices. Thank you **Andreas** and **Jorge** for your availability and modesty. Thank you **Juan Luis** and **Javi** for all the techy discussions, our hacky projects and the affection, to **Solomon** for having "a thing" for hugs and tunics, **Simon** for our debates and your humor. Thank you **Carlo**, **Markus** and **Pavel V** for the hospitality and cheerfulness. **Pablo** for your inclusiveness and charisma. **Alex H** for your habit of waking me up at night asking for

burgers as if my name was McDonald and for being often right and making me learn the hard way. Thank you **Juan Sebastian** for convincing me doing things that I was sure were not "my thing", it ended up making a difference! **Irvin** for our evening gossip and discussions about IT. Thank you **Filip** for our long discussions adventuring in complex and deep topics and for you hospitality. **Alexander S** for your talkativeness and sincerity. **Joakim** and **Antonino** for giving me insight on Swedish-ness; **Frantisek** and **Peter S** for your great music; **Rui** for the good laughs and sharp sarcasm. Thank you **Miguel** and **Javier D** for the movies and games; **Alberto** for your kindness and patience; **Oswaldo**, **Matteo**, **Briac** and **Thomas** for sharing your life experiences. Thank you **Adam**, **Yongjin**, **Tobias** and **Daniel P** for opening my mind with your different views and for your unpredictability. **Michael** for the anime and for redefining my concept of unpredictability. Thank you to the Ycaza-Noboa-Moscoso super-family: **Roberto**, **Emilio**, **Andres**, **Santiago**, **Sebastian** and **Joaquin** for bringing noise, movement and a sincere and affectionate friendliness that is rare to find, stay the way you are!

A special thank you to **Nacho** for being a constant of these seven years in Sweden, thank you for your friendship and hospitality, for the game nights, the exciting scientific explanations and for being of inspiration. Thank you to **Frank** for your friendship, the jokes, the fancy rooftops, golden day in Tokyo and the unforgettable sushi that gave start to our great culinary experiences. A heartfelt thank you **Kadeer**, your esteem and trust in my capabilities and your good soul had a tremendous effect on my future, it really made a big difference. Thank you **Arvin** for listening to my crazy ideas and the dinners. **Liban** for demonstrating to me that the passion for science cannot be neglected for long. **Apostolos** for the explosive cheerfulness. Thank you for **Ester**, **Rodolfo, Elisa** and **Hernan** for teaching me dancing.

Thank you for the friends of a lifetime that when I am back home make me feel as I never left and when I am far away make me feel nostalgic of our Saturdays together: **Claudio**, **Sandra**, **Sofia**, **Riccardo** and **Peppino**. Thank you for all the love and trust and for flying up north to see my graduation.

Ora I ringraziamenti più importanti. Ai miei genitori **Cettina** e **Giuseppe**, senza la vostra comprensione, disponibilità, aiuto, cura e spirito di abnegazione non sarei mai potuto arrivare a realizzare questo sogno. Grazie per avere partecipato ad ogni passo ed avere gioito insieme a me dei miei successi e avermi supportato nei momenti difficili. Grazie a **Sergio** e a **Virginia** per tutto l'affetto, per esserci sempre e per avere compensato la mia assenza. Grazie **Rino**, **Maria Teresa** e **Martina** per tutte le dimostrazioni di affetto. Grazie a **Salvatore** e **Giuseppina** per il grande affetto, per i consigli preoccupati, per essersi sforzati di capire le mie scelte e averne gioito. Grazie ai futuri suoceri e cognata **Fabrizio**, **Giovanna e Alessandra** per l'affetto e a **Ghega** per la predilezione.

Grazie dal profondo del mio cuore a **Marina**, per tutto l'amore, la pazienza, il supporto e soprattutto per avere resistito alla lontananza di tutti questi anni. Grazie per avermi insegnato tanto, per avere condiviso le miei gioie e i successi e per avermi permesso di partecipare alle tue. Per avermi regalato un sorriso con cui andare avanti quando mi sembrava fosse troppo dura. Non potrei essere più felice di cominciare questa grande avventura assieme, ce l'abbiamo fatta!

8 REFERENCES

Alberts, B., Johnson, A., Lewis, J., Morgan, D., Raff, M., Roberts, K., and Walter, P. (2014). Molecular Biology of the Cell 6e.

Alemany, A., Florescu, M., Baron, C.S., Peterson-Maduro, J., and Van Oudenaarden, A. (2018). Whole-organism clone tracing using single-cell sequencing. Nature *556*, 108–112.

Aloisi, F. (2001). Immune function of microglia. Glia *36*, 165–179.

Alon, U. (2007). An introduction to systems biology. Hall/CRC.

Altman, J. (1963). Autoradiographic investigation of cell proliferation in the brains of rats and cats. Anat. Rec. *145*, 573–591.

Altman, J., and Das, G.D. (1966). Autoradiographic and Histological Studies o Postnatal Neurogenesis. J. Comp. Neurol. *126*, 337–389.

Van den Ameele, J., Tiberi, L., Vanderhaeghen, P., and Espuny-Camacho, I. (2014). Thinking out of the dish: What to learn about cortical development using pluripotent stem cells. Trends Neurosci. *37*, 334–342.

Anders, S., and Huber, W. (2010). BIOINFORMATICS Differential expression analysis for sequence count data. Mol. Biol. *100*.

Angevine, J.B. (1965). Time of neuron origin in the hippocampal region. An autoradiographic study in the mouse. Exp. Neurol. Suppl. Suppl 2:1-70.

Anthony, T.E., Klein, C., Fishell, G., and Heintz, N. (2004). Radial glia serve as neuronal progenitors in all regions of the central nervous system. Neuron *41*, 881–890.

Apostolova, G., and Dechant, G. (2009). Development of neurotransmitter phenotypes in sympathetic neurons. Auton. Neurosci. Basic Clin. *151*, 30–38.

Arenas, E., Denham, M., and Villaescusa, J.C. (2015). How to make a midbrain dopaminergic neuron. Development *142*, 1918–1936.

Arendt, D. (2008). The evolution of cell types in animals: emerging principles from molecular studies. Nat. Rev. Genet. *9*, 868–882.

Arendt, D., Musser, J.M., Baker, C.V.H., Bergman, A., Cepko, C., Erwin, D.H., Pavlicev, M., Schlosser, G., Widder, S., Laubichler, M.D., et al. (2016). The origin and evolution of cell types. Nat. Rev. Genet. *17*, 744–757.

Bakken, T.E., Miller, J.A., Ding, S.-L., Sunkin, S.M., Smith, K.A., Ng, L., Szafer, A., Dalley, R.A., Royall, J.J., Lemon, T., et al. (2016). A comprehensive transcriptional map of primate brain development. Nature *535*, 367–375.

Banerjee, A., and Chaudhury, S. (2010). Statistics without tears: Populations and samples. Ind. Psychiatry J. *19*, 60–65.

Banerji, C.R.S., Miranda-Saavedra, D., Severini, S., Widschwendter, M., Enver, T., Zhou, J.X., and Teschendorff, A.E. (2013). Cellular network entropy as the energy potential in Waddington's differentiation landscape. Sci. Rep. *3*, 25–27.

Barker, R.A., Drouin-Ouellet, J., and Parmar, M. (2015). Cell-based therapies for Parkinson disease—past insights and future potential. Nat. Rev. Neurol. *11*, 492–503.

Bendall, S.C., Simonds, E.F., Qiu, P., Amir, E. -a. D., Krutzik, P.O., Finck, R., Bruggner, R. V, Melamed, R., Trejo, A., Ornatsky, O.I., et al. (2011). Single-Cell Mass Cytometry of Differential Immune and Drug Responses Across a Human Hematopoietic Continuum. Science (80-.). *332*, 687–696.

Bendall, S.C., Davis, K.L., Amir, E.A.D., Tadmor, M.D., Simonds, E.F., Chen, T.J., Shenfeld, D.K., Nolan, G.P., and Pe'Er, D. (2014). Single-cell trajectory detection uncovers

progression and regulatory coordination in human b cell development. Cell *157*, 714–725.

Bintu, L., Yong, J., Antebi, Y.E., Mccue, K., Kazuki, Y., Uno, N., Oshimura, M., and Elowitz, M.B. (2016). Supplementary Materials for Dynamics of epigenetic regulation at the single - cell level. *351*.

Birren, S.J., Lo, L., and Anderson, D.J. (1993). Sympathetic neuroblasts undergo a developmental switch in trophic dependence. Development *119*, 597–610.

Blondel, V.D., Guillaume, J.L., Lambiotte, R., and Lefebvre, E. (2008). Fast unfolding of communities in large networks. J. Stat. Mech. Theory Exp. *2008*.

Bonilla, S., Hall, A.C., Pinto, L., Attardo, A., Götz, M., Huttner, W.B., and Arenas, E. (2008). Identification of midbrain floor plate radial glia-like cells as dopaminergic progenitors. Glia *56*, 809–820.

Brady, G., Barbara, M., and Iscove, N.N. (1990). Representative in vitro cDNA amplification from individual hemopoietic cells and colonies. Methods Mol. Cell. Biol. *2*, 17–25.

Brody, T., and Odenwald, W.F. (2005). Regulation of temporal identities during Drosophila neuroblast lineage development. Curr. Opin. Cell Biol. *17*, 672–675.

Brunet, T., Fischer, A.H.L., Steinmetz, P.R.H., Lauri, A., Bertucci, P., and Arendt, D. (2016). The evolutionary origin of bilaterian smooth and striated myocytes. Elife *5*, 1–24.

Buettner, F., Pratanwanich, N., McCarthy, D.J., Marioni, J.C., and Stegle, O. (2017). f-scLVM: Scalable and versatile factor analysis for single-cell RNA-seq. Genome Biol. *18*, 1–13.

Butler, A., Hoffman, P., Smibert, P., Papalexi, E., and Satija, R. (2018). Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nat. Biotechnol. *36*, 411–420.

Cai, X., Evrony, G.D., Lehmann, H.S., Elhosary, P.C., Mehta, B.K., Poduri, A., and Walsh, C.A. (2014). Single-Cell, Genome-wide Sequencing Identifies Clonal Somatic Copy-Number Variation in the Human Brain. Cell Rep. *8*, 1280–1289.

Cameron, H.A., and Mckay, R.D.G. (2001). Adult neurogenesis produces a large pool of new granule cells in the dentate gyrus. J. Comp. Neurol. *435*, 406–417.

Cao, J., Packer, J.S., Ramani, V., Cusanovich, D.A., Huynh, C., Daza, R., Qiu, X., Lee, C., Furlan, S.N., Steemers, F.J., et al. (2017). Comprehensive single-cell transcriptional profiling of a multicellular organism. Science (80-.). *357*, 661–667.

Chatton, É.P.L. (1938). Titres et travaux scientifiques (1906-1937) de Edouard Chatton (Impr. E. Sottano).

Chen, T., He, H., and Church, G. (1999). Modeling Gene Expression With Differential Equations. Pacific Symp. … *2*, 29–40.

Chung, C.Y., Seo, H., Sonntag, K.C., Brooks, A., Lin, L., and Isacson, O. (2005). Cell type-specific gene expression of midbrain dopaminergic neurons reveals molecules involved in their vulnerability and protection. Hum. Mol. Genet. *14*, 1709–1725.

Chung, S., Leung, A., Han, B.S., Chang, M.Y., Moon, J. Il, Kim, C.H., Hong, S., Pruszak, J., Isacson, O., and Kim, K.S. (2009). Wnt1-lmx1a Forms a Novel Autoregulatory Loop and Controls Midbrain Dopaminergic Differentiation Synergistically with the SHH-FoxA2 Pathway. Cell Stem Cell *5*, 646–658.

Clevers, H., Rafelski, S., Elowitz, M., and Lein, E. (2017). What Is Your Conceptual Definition of "Cell Type" in the Context of a Mature Organism? Cell Syst. *4*, 255–259.

Collart, C., Allen, G.E., Bradshaw, C.R., Smith, J.C., and Zegerman, P. (2013). Titration of Four Replication Factors Is Essential for the Xenopus laevis Midblastula Transition. Science (80-.). *341*, 893–896.

Damier, P., Hirsch, E.C., Agid, Y., and Graybiel, A.M. (1999). The substantia nigra of the human brain. I. Nigrosomes and the nigral matrix, a compartmental organization based on calbindin D(28K) immunohistochemistry. Brain *122 (Pt 8*, 1421–1436.

Davis-Dusenbery, B.N., Williams, L.A., Klim, J.R., and Eggan, K. (2014). How to make spinal motor neurons. Development *141*, 491–501.

Deans, C., and Maggert, K.A. (2015). What do you mean, “Epigenetic”? Genetics *199*, 887–896.

DeFelipe, J., López-Cruz, P.L., Benavides-Piccione, R., Bielza, C., Larrañaga, P., Anderson, S., Burkhalter, A., Cauli, B., Fairén, A., Feldmeyer, D., et al. (2013). New insights into the classification and nomenclature of cortical GABAergic interneurons. Nat. Rev. Neurosci. *14*, 202–216.

Diaconis, P., Goel, S., and Holmes, S. (2008). Horseshoes in multidimensional scaling and local kernel methods. Ann. Appl. Stat. *2*, 777–807.

van Dijk, D., Sharma, R., Nainys, J., Yim, K., Kathail, P., Carr, A.J., Burdziak, C., Moon, K.R., Chaffer, C.L., Pattabiraman, D., et al. (2018). Recovering Gene Interactions from Single-Cell Data Using Data Diffusion. Cell *174*, 716–729.e27.

Eagleson, G.W., and Harris, W.A. (1990). Mapping of the presumptive brain regions in the neural plate ofXenopus laevis. J. Neurobiol. *21*, 427–440.

Eberwine, J., Yeh, H., Miyashiro, K., Cao, Y., Nair, S., Finnell, R., Zettel, M., and Coleman, P. (1992). Analysis of gene expression in single live neurons. Proc. Natl. Acad. Sci. U. S. A. *89*, 3010–3014.

Ebisuya, M., and Briscoe, J. (2018). What does time mean in development ? Development *145*, 1–6.

Elowitz, M.B., Levine, A.J., Siggia, E.D., and Swain, P.S. (2002). Stochastic gene expression in a single cell. Science (80-.). *297*, 1183–1186.

Ester, M., Xu, X., Kriegel, H., and Sander, J. (1996). Density-based algorithm for discovering clusters in large spatial databases with noise. Proc. Acm Sigkdd Int. Conf. Knowl. Discov. Data Min. *pages*, 226–231.

Fan, H.C., Fu, G.K., and Fodor, S.P. a. (2015). Combinatorial labeling of single cells for gene expression cytometry. Science (80-.). *347*, 1258367–1258367.

Feil, R., Metzger, D., and Chambon, P. (1997). Regulation of Cre Recombinase Activity by Mutated Estrogen Receptor Ligand-Binding Domains. *757*, 752–757.

Finak, G., McDavid, A., Yajima, M., Deng, J., Gersuk, V., Shalek, A.K., Slichter, C.K., Miller, H.W., McElrath, M.J., 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.

Fishell, G., and Rudy, B. (2011). Mechanisms of inhibition within the telencephalon: “where the wild things are”.

Fortunato, S., and Barthélemy, M. (2007). Resolution limit in community detection. Pnas *104*, 36–41.

Frieda, K.L., Linton, J.M., Hormoz, S., Choi, J., Chow, K.H.K., Singer, Z.S., Budde, M.W., Elowitz, M.B., and Cai, L. (2017). Synthetic recording and in situ readout of lineage information in single cells. Nature *541*, 107–111.

Fu, Y.H., Yuan, Y., Halliday, G., Rusznák, Z., Watson, C., and Paxinos, G. (2012). A cytoarchitectonic and chemoarchitectonic analysis of the dopamine cell groups in the substantia nigra, ventral tegmental area, and retrorubral field in the mouse. Brain Struct. Funct. *217*, 591–612.

Furlan, A., La Manno, G., Lübke, M., Häring, M., Abdo, H., Hochgerner, H., Kupari, J., Usoskin, D., Airaksinen, M.S., Oliver, G., et al. (2016). Visceral motor neuron diversity delineates a cellular basis for nipple- and pilo-erection muscle control. Nat. Neurosci. *19*, 1331–1340.

Gardner, T.S., Bernardo, D., Lorenz, D., and Collins, J.J. (2015). Inferring Genetic Networks and Identifying Compound Mode of Action via Expression Profiling Author(s): Timothy S. Gardner, Diego di Bernardo, David Lorenz and James J. Collins Source: *301*, 102–105.

Gebhardt, R. (1992). Metabolic zonation of the liver: Regulation and implications for liver function. Pharmacol. Ther. *53*, 275–354.

Van Gelder, R.N., von Zastrow, M.E., Yool, a, Dement, W.C., Barchas, J.D., and Eberwine, J.H. (1990). Amplified RNA synthesized from limited quantities of heterogeneous cDNA. Proc. Natl. Acad. Sci. U. S. A. *87*, 1663–1667.

Gibb, S., Maroto, M., and Dale, J.K. (2010). The segmentation clock mechanism moves up a notch. Trends Cell Biol. *20*, 593–600.

Giladi, A., and Amit, I. (2018). Single-Cell Genomics: A Stepping Stone for Future Immunology Discoveries. Cell *172*, 14–21.

Gillespie, D.T. (1977). Exact stochastic simulation of coupled chemical reactions. J. Phys. Chem. *81*, 2340–2361.

Ginhoux, F., Greter, M., Leboeuf, M., Nandi, S., See, P., Gokhan, S., Mehler, M.F., Conway, S.J., Ng, L.G., Stanley, E.R., et al. (2010). Fate Mapping Analysis Reveals That Adult Microglia Derive from Primitive Macrophages. Science (80-.). *330*, 841–845.

Gokce, O., Stanley, G.M., Treutlein, B., Neff, N.F., Camp, J.G., Malenka, R.C., Rothwell, P.E., Fuccillo, M. V., S??dhof, T.C., and Quake, S.R. (2016). Cellular Taxonomy of the Mouse Striatum as Revealed by Single-Cell RNA-Seq. Cell Rep. *16*, 1126–1137.

Goldmann, T., Wieghofer, P., Jordão, M.J.C., 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.

Gonzalez-Lozano, M.A., Klemmer, P., Gebuis, T., Hassan, C., Van Nierop, P., Van Kesteren, R.E., Smit, A.B., and Li, K.W. (2016). Dynamics of the mouse brain cortical synaptic proteome during postnatal brain development. Sci. Rep. *6*, 1–15.

Goridis, C., and Rohrer, H. (2002). Specification of catecholaminergic and serotonergic neurons. Nat. Rev. Neurosci. *3*, 531–541.

Graham, A. (2003). The neural crest. Curr. Biol. *13*, R381–R384.

Grove, E.A., and Monuki, E.S. (2013). Morphogens, Patterning Centers, and their Mechanisms of Action. In Patterning and Cell Type Specification in the Developing CNS and PNS, (Elsevier), pp. 25–44.

Grove, E.A., Kirkwood, T.B.L., and Price, J. (1992). Neuronal precursor cells in the rat hippocampal formation contribute to more than one cytoarchitectonic area. Neuron *8*, 217–229.

Grün, D., Kester, L., and van Oudenaarden, A. (2014). Validation of noise models for single-cell transcriptomics. Nat. Methods *11*, 637–640.

Grün, D., Lyubimova, A., Kester, L., Wiebrands, K., Basak, O., Sasaki, N., Clevers, H., and Van Oudenaarden, A. (2015). Single-cell messenger RNA sequencing reveals rare intestinal cell types. Nature *525*, 251–255.

Guillemot, F., and Joyner, A.L. (1993). Dynamic expression of the murine Achaete-Scute homologue Mash-1 in the developing nervous system. Mech. Dev. *42*, 171–185.

Haghverdi, L., Lun, A.T.L., Morgan, M.D., and Marioni, J.C. (2018). Batch effects in single-

cell RNA-sequencing data are corrected by matching mutual nearest neighbors. Nat. Biotechnol. *36*, 421–427.

Hall, B.K. (2007). Keywords and concepts in evolutionary developmental biology (Discovery Publishing House).

Halpern, K.B., Shenhav, R., Matcovitch-Natan, O., Tóth, B., Lemze, D., Golan, M., Massasa, E.E., Baydatch, S., Landen, S., Moor, A.E., et al. (2017). Single-cell spatial reconstruction reveals global division of labour in the mammalian liver. Nature *542*, 1–5.

Hari, L., Brault, V., Kléber, M., Lee, H.Y., Ille, F., Leimeroth, R., Paratore, C., Suter, U., Kemler, R., and Sommer, L. (2002). Lineage-specific requirements of β-catenin in neural crest development. J. Cell Biol. *159*, 867–880.

Harland, R., and Gerhart, J. (1997). Formation and Function of Spemann ' S Organizer. Annu. Rev. Cell. andDevelopmental Biol. *13*, 611–667.

Harrison, D.A., and Perrimon, N. (1993). Simple and efficient generation of marked clones in Drosophila. Curr. Biol. *3*, 424–433.

Hashimshony, T., Wagner, F., Sher, N., and Yanai, I. (2012). CEL-Seq: Single-Cell RNA-Seq by Multiplexed Linear Amplification. Cell Rep. *2*, 666–673.

Hilfinger, A., Norman, T.M., and Paulsson, J. (2016). Exploiting Natural Fluctuations to Identify Kinetic Mechanisms in Sparsely Characterized Systems. Cell Syst. *2*, 251–259.

Hochgerner, H., Zeisel, A., Lönnerberg, P., and Linnarsson, S. (2018). Conserved properties of dentate gyrus neurogenesis across postnatal development revealed by single-cell RNA sequencing. Nat. Neurosci. *21*, 290–299.

Holdorff, B. (2002). Friedrich Heinrich Lewy (1885-1950) and his work. J Hist Neurosci *11*, 19–28.

Holliday, R. (1994). Epigenetics : An Overview. *453457*.

Hoyle, N.P., and Ish-Horowicz, D. (2013). Transcript processing and export kinetics are rate-limiting steps in expressing vertebrate segmentation clock genes. Proc. Natl. Acad. Sci. *110*, E4316–E4324.

Huang, M., Wang, J., Torre, E., Dueck, H., Shaffer, S., Bonasio, R., Murray, J.I., Raj, A., Li, M., and Zhang, N.R. (2018). SAVER: Gene expression recovery for single-cell RNA sequencing. Nat. Methods *15*, 539–542.

Hubaud, A., and Pourquié, O. (2014). Signalling dynamics in vertebrate segmentation. Nat. Rev. Mol. Cell Biol. *15*, 709–721.

Iacono, G., Mereu, E., Guillaumet-adkins, A., Corominas, R., Cuscó, I., Rodríguez-esteban, G., Gut, M., Pérez-jurado, L.A., Gut, I., and Heyn, H. (2018). bigSCale : an analytical framework for big-scale single-cell data. 878–890.

Imayoshi, I., Sakamoto, M., Ohtsuka, T., Takao, K., Miyakawa, T., Yamaguchi, M., Mori, K., Ikeda, T., Itohara, S., and Kageyama, R. (2008). Roles of continuous neurogenesis in the structural and functional integrity of the adult forebrain. Nat. Neurosci. *11*, 1153–1161.

Islam, S., Kjällquist, U., Moliner, A., Zajac, P., Fan, J.B., Lönnerberg, P., and Linnarsson, S. (2011). Characterization of the single-cell transcriptional landscape by highly multiplex RNA-seq. Genome Res. *21*, 1160–1167.

Islam, S., Zeisel, A., Joost, S., La Manno, G., Zajac, P., Kasper, M., Lönnerberg, P., and Linnarsson, S. (2013). Quantitative single-cell RNA-seq with unique molecular identifiers. Nat. Methods *11*, 163–166.

Islam, S., Zeisel, A., Joost, S., La Manno, G., Zajac, P., Kasper, M., L??nnerberg, P., and Linnarsson, S. (2014). Quantitative single-cell RNA-seq with unique molecular identifiers. Nat. Methods *11*, 163–166.

Jacobson, M., and Hirose, G. (1978). Origin of the retina from both sides of the embryonic brain: A contribution to the problem of crossing at the optic chiasma. Science (80-.). *202*, 637–639.

Jaitin, D.A., Weiner, A., Yofe, I., Lara-Astiaso, D., Keren-Shaul, H., David, E., Salame, T.M., Tanay, A., van Oudenaarden, A., and Amit, I. (2016). Dissecting Immune Circuits by Linking CRISPR-Pooled Screens with Single-Cell RNA-Seq. Cell *167*, 1883–1896.e15.

Jessen, K.R., and Mirsky, R. (2005). The origin and development of glial cells in peripheral nerves. Nat. Rev. Neurosci. *6*, 671.

Johnson, W.E., Li, C., and Rabinovic, A. (2007). Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics *8*, 118–127.

Joost, S., Zeisel, A., Jacob, T., Sun, X., La Manno, G., L??nnerberg, P., Linnarsson, S., and Kasper, M. (2016). Single-Cell Transcriptomics Reveals that Differentiation and Spatial Signatures Shape Epidermal and Hair Follicle Heterogeneity. Cell Syst. *3*, 221–237.e9.

Jungermann, K., and Keitzmann, T. (1996). Zonation of Parenchymal and Nonparenchymal Metabolism in Liver. Annu. Rev. Nutr. *16*, 179–203.

Kalhor, R., Mali, P., and Church, G.M. (2017). Rapidly evolving homing CRISPR barcodes. Nat. Methods *14*, 195–200.

Kalhor, R., Kalhor, K., Leeper, K., Graveline, A., Mali, P., and Church, G.M. (2018). A homing CRISPR mouse resource for barcoding and lineage tracing. bioRxiv 280289.

Kandel, E.R., Schwartz, J.H., Jessell, T.M., A, S.S., and J, H.A. (2013). Principles of Neural Science.

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

Kester, L., and van Oudenaarden, A. (2018). Single-Cell Transcriptomics Meets Lineage Tracing. Cell Stem Cell 1–14.

Kharchenko, P. V, Silberstein, L., and Scadden, D.T. (2014). Bayesian approach to single-cell differential expression analysis. Nat. Methods *11*, 740–742.

Kiecker, C., and Lumsden, A. (2012). The Role of Organizers in Patterning the Nervous System. Annu. Rev. Neurosci. *35*, 347–367.

Kimmel, C.B., Warga, R.M., and Schilling, T.F. (1990). Origin and organization of the zebrafish fate map. Development *108*, 581–594.

Kivioja, T., Vähärautio, A., Karlsson, K., and Bonke, M. (2011). Counting absolute numbers of molecules using unique molecular identifiers. Nat. … *9*, 1–5.

Klausberger, T., and Somogyi, P. (2008). Neuronal diversity and temporal dynamics: the unity of hippocampal circuit operations. Science *321*, 53–57.

Klein, A.M.M., Mazutis, L., Akartuna, I., Tallapragada, N., Veres, A., Li, V., Peshkin, L., Weitz, D.A.A., and Kirschner, M.W.W. (2015). Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. Cell *161*, 1187–1201.

Kretzschmar, K., and Watt, F.M. (2012). Lineage tracing. Cell *148*, 33–45.

Kriegstein, A., and Alvarez-Buylla, A. (2009). The Glial Nature of Embryonic and Adult Neural Stem Cells. Annu. Rev. Neurosci. *32*, 149–184.

Kurimoto, K., Yabuta, Y., Ohinata, Y., Ono, Y., Uno, K.D., Yamada, R.G., Ueda, H.R., and Saitou, M. (2006). An improved single-cell cDNA amplification method for efficient high-density oligonucleotide microarray analysis. Nucleic Acids Res. *34*.

L. Lun, A.T., Bach, K., and Marioni, J.C. (2016). Pooling across cells to normalize single-cell RNA sequencing data with many zero counts. Genome Biol. *17*, 75.

Lancichinetti, A., and Fortunato, S. (2009). Community detection algorithms: A comparative analysis. Phys. Rev. E - Stat. Nonlinear, Soft Matter Phys. *80*, 1–11.

Lancichinetti, A., and Fortunato, S. (2011). Limits of modularity maximization in community detection. Phys. Rev. E - Stat. Nonlinear, Soft Matter Phys. *84*, 1–8.

Landgraf, D., Koch, C.E., and Oster, H. (2014). Embryonic development of circadian clocks in the mammalian suprachiasmatic nuclei. Front. Neuroanat. *8*, 1–7.

Lee, D.D., and Seung, H.S. (1999). Learning the parts of objects by non-negative matrix factorization. Nature *401*, 788–791.

de Leeuw, J., and Leeuw, J. a N.D.E. (2007). A horseshoe for multidimensional scaling.

Lei, J. (2011). Stochastic Modeling in Systems Biology. 1–25.

Lein, E.S., Hawrylycz, M.J., Ao, N., Ayres, M., Bensinger, A., Bernard, A., Boe, A.F., Boguski, M.S., Brockway, K.S., Byrnes, E.J., et al. (2007). Genome-wide atlas of gene expression in the adult mouse brain. Nature *445*, 168–176.

Leone, D.P., Srinivasan, K., Chen, B., Alcamo, E., and McConnell, S.K. (2008). The determination of projection neuron identity in the developing cerebral cortex. Curr. Opin. Neurobiol. *18*, 28–35.

Li, W.V., and Li, J.J. (2018). An accurate and robust imputation method scImpute for single-cell RNA-seq data. Nat. Commun. *9*, 1–9.

Li, C., Cesbron, F., Oehler, M., Brunner, M., and Höfer, T. (2018). Frequency Modulation of Transcriptional Bursting Enables Sensitive and Rapid Gene Regulation. Cell Syst. *6*, 409–423.e11.

Li, X., Brock, G.N., Rouchka, E.C., Cooper, N.G.F., Wu, D., O'Toole, T.E., Gill, R.S., Eteleeb, A.M., O'Brien, L., and Rai, S.N. (2017). A comparison of per sample global scaling and per gene normalization methods for differential expression analysis of RNA-seq data. PLoS One *12*, e0176185.

Lin, P., Troup, M., and Ho, J.W.K. (2017). CIDR: Ultrafast and accurate clustering through imputation for single-cell RNA-seq data. Genome Biol. *18*, 1–11.

Lindvall, O., and Kokaia, Z. (2009). Prospects of stem cell therapy for replacing dopamine neurons in Parkinson's disease. Trends Pharmacol. Sci. *30*, 260–267.

Lodato, M.A., Woodworth, M.B., Lee, S., Evrony, G.D., Mehta, B.K., Karger, A., Lee, S., Chittenden, T.W., D'Gama, A.M., Cai, X., et al. (2015). Somatic mutation in single human neurons tracks developmental and transcriptional history. Science (80-.). *350*, 94–98.

Lois, C., and Alarez-Buylla, A. (1994). Long-distance neuronal migration in the adult mammalian brain. Science (80-.). *264*, 1145–1148.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 1–21.

der Maaten, L., and Hinton, G. (2008). Visualizing data using t-SNE. J. Mach. Learn. Res. *9*, 85.

MacDonald, J.L., Fame, R.M., Azim, E., Shnider, S.J., Molyneaux, B.J., Arlotta, P., and Macklis, J.D. (2013). Specification of Cortical Projection Neurons. In Patterning and Cell Type Specification in the Developing CNS and PNS, (Elsevier), pp. 475–502.

Macosko, E.Z., Basu, A., Satija, R., Nemesh, J., Shekhar, K., Goldman, M., Tirosh, I., Bialas, A.R., Kamitaki, N., Martersteck, E.M., et al. (2015a). Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. Cell *161*, 1202–1214.

Macosko, E.Z., Basu, A., Satija, R., Nemesh, J., Shekhar, K., Goldman, M., Tirosh, I., Bialas, A.R., Kamitaki, N., Martersteck, E.M., et al. (2015b). Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. Cell *161*, 1202–1214.

Magistretti, P.J. (2011). Neuron-glia metabolic coupling and plasticity. Exp Physiol *96*, 407–410.

Malatesta, P., Hack, M.A., Hartfuss, E., Kettenmann, H., Klinkert, W., Kirchhoff, F., and Götz, M. (2003). Neuronal or glial progeny: Regional differences in radial glia fate. Neuron *37*, 751–764.

La Manno, G., Gyllborg, D., Codeluppi, S., Nishimura, K., Salto, C., Zeisel, A., Borm, L.E., Stott, S.R.W., Toledo, E.M., Villaescusa, J.C., et al. (2016). Molecular Diversity of Midbrain Development in Mouse, Human, and Stem Cells. Cell *167*, 566–580.e19.

Marioni, J.C., and Arendt, D. (2017). How Single-Cell Genomics Is Changing Evolutionary and Developmental Biology.

Marioni, J.C., Mason, C.E., Mane, S.M., Stephens, M., and Gilad, Y. (2008). comparison with gene expression arrays RNA-seq : An assessment of technical reproducibility and comparison with gene expression arrays. 1509–1517.

Marques, S., Zeisel, A., Codeluppi, S., van Bruggen, D., Mendanha Falcao, A., Xiao, L., Li, H., Haring, M., Hochgerner, H., Romanov, R.A., et al. (2016). Oligodendrocyte heterogeneity in the mouse juvenile and adult central nervous system. Science (80-.). *352*, 1326–1329.

McInnes, L., and Healy, J. (2018). UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. 1–18.

Mckenna, A., Findlay, G.M., Gagnon, J.A., Horwitz, M.S., Schier, A.F., and Shendure, J. (2016). Whole organism lineage tracing by combinatorial and cumulative genome editing HHS Public Access. Sci. July *29*, 1–27.

Milholland, B., Dong, X., Zhang, L., Hao, X., Suh, Y., and Vijg, J. (2017). Differences between germline and somatic mutation rates in humans and mice. Nat. Commun. *8*, 1–8.

Molyneaux, B.J., Arlotta, P., Menezes, J.R.L., and Macklis, J.D. (2007). Neuronal subtype specification in the cerebral cortex. Nat. Rev. Neurosci. *8*, 427–437.

Moody, S.A. (1987a). Fates of the blastomeres of the 32-cell-stage Xenopus embryo. Dev. Biol. *122*, 300–319.

Moody, S.A. (1987b). Fates of the blastomeres of the 16-cell stage Xenopus embryo. Dev. Biol. *119*, 560–578.

Mora-Bermúdez, F., Badsha, F., Kanton, S., Camp, J.G., Vernot, B., Köhler, K., Voigt, B., Okita, K., Maricic, T., He, Z., et al. (2016). Differences and similarities between human and chimpanzee neural progenitors during cerebral cortex development. Elife *5*, 1–24.

Musumeci, G. (2014). Past, present and future: overview on histology and histopathology. J. Histol. Histopathol. *1*, 5.

Nakamura, H. (2013). Midbrain Patterning. In Patterning and Cell Type Specification in the Developing CNS and PNS, (Elsevier), pp. 45–60.

O'Keefe, J. (1976). Place units in the hippocampus of the freely moving rat. Exp. Neurol. *51*, 78–109.

O'Leary, D.D.M., Stocker, A.M., and Zembrzycki, A. (2013). Area Patterning of the Mammalian Cortex. Compr. Dev. Neurosci. Patterning Cell Type Specif. Dev. CNS PNS 61–85.

Olimpio, E.P., Dang, Y., and Youk, H. (2017). Statistical dynamics of spatial-order formation by communicating cells. ISCIENCE *2*, 27–40.

Otani, T., Marchetto, M.C., Gage, F.H., Simons, B.D., and Livesey, F.J. (2016). 2D and 3D Stem Cell Models of Primate Cortical Development Identify Species-Specific Differences in Progenitor Behavior Contributing to Brain Size. Cell Stem Cell *18*, 467–480.

Owen, R. (1848). On the Archetype and Homologies of the Vertebrate Skeleton (Van Voorst).

Parkinson, J. (2002). An essay on the shaking palsy. 1817. J. Neuropsychiatry Clin. Neurosci. *14*, 223–36; discussion 222.

Paul, F., Arkin, Y., Giladi, A., Jaitin, D.A., Kenigsberg, E., Keren-Shaul, H., Winter, D., Lara-Astiaso, D., Gury, M., Weiner, A., et al. (2015). Transcriptional Heterogeneity and Lineage Commitment in Myeloid Progenitors. Cell *163*, 1663–1677.

Perli, S.D., Cui, C.H., and Lu, T.K. (2016). Continuous genetic recording with self-targeting CRISPR-Cas in human cells. *353*.

Picelli, S., Björklund, Å.K., Faridani, O.R., Sagasser, S., Winberg, G., and Sandberg, R. (2013). Smart-seq2 for sensitive full-length transcriptome profiling in single cells. Nat. Methods *10*, 1096–1098.

Poulin, J.F., Zou, J., Drouin-Ouellet, J., Kim, K.Y.A., Cicchetti, F., and Awatramani, R.B. (2014). Defining midbrain dopaminergic neuron diversity by single-cell gene expression profiling. Cell Rep. *9*, 930–943.

Prakash, N., Puelles, E., Freude, K., Trümbach, D., Omodei, D., Di Salvio, M., Sussel, L., Ericson, J., Sander, M., Simeone, A., et al. (2009). Nkx6-1 controls the identity and fate of red nucleus and oculomotor neurons in the mouse midbrain. Development *136*, 2545–2555.

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

Qian, X., Shen, Q., Goderie, S.K., He, W., Capela, A., Davis, A.A., and Temple, S. (2000). Timing of CNS Cell Generation. Neuron *28*, 69–80.

Quiroga, Q., Reddy, L., Kreiman, G., Koch, C., and Fried, I. (2005). Invariant visual representation by single neurons in the human brain. Nature *435*, 1102–1107.

Raj, A., and van Oudenaarden, A. (2008). Nature, Nurture, or Chance: Stochastic Gene Expression and Its Consequences. Cell *135*, 216–226.

Raj, B., Wagner, D.E., McKenna, A., Pandey, S., Klein, A.M., Shendure, J., Gagnon, J.A., and Schier, A.F. (2018). Simultaneous single-cell profiling of lineages and cell types in the vertebrate brain. Nat. Biotechnol. *36*, 442–450.

Ramón Cajal, S. (1904). Textura del Sistema Nervioso del Hombre y de los Vertebrados (Madrid Nicolas Moya).

Regev, A., Teichmann, S.A., Lander, E.S., Amit, I., Benoist, C., Birney, E., Bodenmiller, B., Campbell, P., Carninci, P., Clatworthy, M., et al. (2017). The human cell atlas. Elife *6*, 1–30.

Reissmann, E., Ernsberger, U., Francis-West, P.H., Rueger, D., Brickell, P.M., and Rohrer, H. (1996). Involvement of bone morphogenetic protein-4 and bone morphogenetic protein-7 in the differentiation of the adrenergic phenotype in developing sympathetic neurons. Development *122*, 2079–2088.

Remak, R. (1852). Ueber extracellulare Entstehung thierischer Zellen und über Vermehrung derselben durch Theilung. Müllers Arch. Anat. Physiol *1852*, 47–57.

Risso, D., Ngai, J., Speed, T.P., and Dudoit, S. (2014). Normalization of RNA-seq data using factor analysis of control genes or samples. Nat. Biotechnol. *32*.

Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). Limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. *43*, e47.

Roeper, J. (2013). Dissecting the diversity of midbrain dopamine neurons. Trends Neurosci. *36*, 336–342.

Romanov, R.A., Zeisel, A., Bakker, J., Girach, F., Hellysaz, A., Tomer, R., Alpár, A.,

Mulder, J., Clotman, F., Keimpema, E., et al. (2016). Molecular interrogation of hypothalamic organization reveals distinct dopamine neuronal subtypes. Nat. Neurosci. *20*.

Rosenberg, A.B., Roco, C.M., Muscat, R.A., Kuchina, A., Sample, P., Yao, Z., Gray, L., Peeler, D.J., Mukherjee, S., Chen, W., et al. (2018). Single-cell profiling of the developing mouse brain and spinal cord with split-pool barcoding. Science (80-.). *12*, eaam8999.

Saelens, W., Cannoodt, R., Todorov, H., and Saeys, Y. (2018). A comparison of single-cell trajectory inference methods: towards more accurate and robust tools. bioRxiv 276907.

Salvador-Martínez, I., Grillo, M., Averof, M., and Telford, M.J. (2018). Is it possible to reconstruct an accurate cell lineage using CRISPR recorders? 1–21.

Satija, R., Farrell, J. a, Gennert, D., Schier, A.F., and Regev, A. (2015). Spatial reconstruction of single-cell gene expression data. Nat. Biotechnol.

Saunders, A., Macosko, E., Wysoker, A., Goldman, M., Krienen, F., Rivera, H. de, Bien, E., Baum, M., Wang, S., Goeva, A., et al. (2018). A Single-Cell Atlas of Cell Types, States, and Other Transcriptional Patterns from Nine Regions of the Adult Mouse Brain. bioRxiv 299081.

Schiebinger, G., Shu, J., Tabaka, M., Cleary, B., Subramanian, V., Solomon, A., Liu, S., Lin, S., Berube, P., Lee, L., et al. (2017). Reconstruction of developmental landscapes by optimal-transport analysis of single-cell gene expression sheds light on cellular reprogramming. bioRxiv (Preprint) 191056.

Schleiden, M.J., Schwann, T., and Smith, H. (1847). Microscopical researches into the accordance in the structure and growth of animals and plants. Translated from the German of Theodor Schwann, by Henry Smith. (London,: Sydenham Society,).

Schneider, C., Wicht, H., Enderich, J., Wegner, M., and Rohrer, H. (1999). Bone morphogenetic proteins are required in vivo for the generation of sympathetic neurons. Neuron *24*, 861–870.

Schwann, T.H. (1847). Microscopical Researches into the Accordance in the Structure and Growth of Animals and Plants (Рипол Классик).

Sebé-Pedrós, A., Saudemont, B., Chomsky, E., Plessier, F., Mailhé, M.P., Renno, J., Loe-Mie, Y., Lifshitz, A., Mukamel, Z., Schmutz, S., et al. (2018a). Cnidarian Cell Type Diversity and Regulation Revealed by Whole-Organism Single-Cell RNA-Seq. Cell *173*, 1520–1534.e20.

Sebé-Pedrós, A., Chomsky, E., Pang, K., Lara-Astiaso, D., Gaiti, F., Mukamel, Z., Amit, I., Hejnol, A., Degnan, B.M., and Tanay, A. (2018b). Early metazoan cell type diversity and the evolution of multicellular gene regulation. Nat. Ecol. Evol.

Setty, M., Tadmor, M.D., Reich-Zeliger, S., Angel, O., Salame, T.M., Kathail, P., Choi, K., Bendall, S., Friedman, N., and Pe'Er, D. (2016). Wishbone identifies bifurcating developmental trajectories from single-cell data. Nat. Biotechnol. *34*, 637–645.

Shapiro, E. (2018). On the journey from nematode to human, scientists dive by the zebrafish cell lineage tree. Genome Biol. *19*, 18–20.

Shapiro, E., Biezuner, T., and Linnarsson, S. (2013). Single-cell sequencing-based technologies will revolutionize whole-organism science. Nat. Rev. Genet. *14*, 618–630.

Shen, Q., Wang, Y., Dimos, J.T., Fasano, C.A., Phoenix, T.N., Lemischka, I.R., Ivanova, N.B., Stifani, S., Morrisey, E.E., and Temple, S. (2006). The timing of cortical neurogenesis is encoded within lineages of individual progenitor cells. Nat. Neurosci. *9*, 743–751.

Shumyatsky, G.P., Tsvetkov, E., Malleret, G., Vronskaya, S., Hatton, M., Hampton, L., Battey, J.F., Dulac, C., Kandel, E.R., and Bolshakov, V.Y. (2002). Identification of a signaling network in lateral nucleus of amygdala important for inhibiting memory specifically related to learned fear. Cell *111*, 905–918.

Somogyi, P., and Klausberger, T. (2005). Defined types of cortical interneurone structure space and spike timing in the hippocampus. J. Physiol. *562*, 9–26.

Spanjaard, B., Hu, B., Mitic, N., Olivares-Chauvet, P., Janjuha, S., Ninov, N., and Junker, J.P. (2018). Simultaneous lineage tracing and cell-type identification using CrIsPr-Cas9-induced genetic scars. Nat. Biotechnol. *36*, 469–473.

Strogatz, S.H. (1994). Nonlinear Dynamics and Chaos. Library (Lond). *48*, 498.

Stumpf, M.P.H., Balding, D.J., and Girolami, M. (2011). Handbook of Statistical Systems Biology.

Sugino, K., Hempel, C.M., Miller, M.N., Hattox, A.M., Shapiro, P., Wu, C., Huang, Z.J., and Nelson, S.B. (2006). Molecular taxonomy of major neuronal classes in the adult mouse forebrain. Nat. Neurosci. *9*, 99–107.

Sulston, J.E., Schierenberg, E., White, J.G., and Thomson, J.N. (1983). The embryonic cell lineage of the nematode Caenorhabditis elegans. Dev. Biol. *100*, 64–119.

Tanabe, Y., William, C., and Jessell, T.M. (1998). Specification of Motor Neuron Identity by the MNR2 Homeodomain Protein. Cell *95*, 67–80.

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

Tasic, B., Menon, V., Nguyen, T.N.T., Kim, T.T.K., Jarsky, T., Yao, Z., Levi, B.B., Gray, L.T., Sorensen, S.A., Dolbeare, T., et al. (2016). Adult mouse cortical cell taxonomy revealed by single cell transcriptomics. Nat. Neurosci. *advance on*, 1–37.

Theveneau, E., and Mayor, R. (2012). Neural crest delamination and migration: From epithelium-to-mesenchyme transition to collective cell migration. Dev. Biol. *366*, 34–54.

Tosches, M.A., Yamawaki, T.M., Naumann, R.K., Jacobi, A.A., Tushev, G., and Laurent, G. (2018). Evolution of pallium, hippocampus, and cortical cell types revealed by single-cell transcriptomics in reptiles. Science (80-.). *360*, 881–888.

Trapnell, C. (2015). Defining cell types and states with single-cell genomics. Genome Res. *25*, 1491–1498.

Trapnell, C., Cacchiarelli, D., Grimsby, J., Pokharel, P., Li, S., Morse, M., Lennon, N.J., Livak, K.J., Mikkelsen, T.S., and Rinn, J.L. (2014). The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. Nat. Biotechnol. *32*, 381–386.

Treutlein, B., Brownfield, D.G., Wu, A.R., Neff, N.F., Mantalas, G.L., Espinoza, F.H., Desai, T.J., Krasnow, M.A., and Quake, S.R. (2014). Reconstructing lineage hierarchies of the distal lung epithelium using single-cell RNA-seq. Nature *509*, 371–375.

Tsacopoulos, M., Poitry-Yamate, C.L., Poitry, S., Perrottet, P., and Veuthey, a L. (1997). The nutritive function of glia is regulated by signals released by neurons. Glia *21*, 84–91.

Turner, D.L., and Cepko, C.L. (1987). A common progenitor for neurons and glia persists in rat retina late in development. Nature *328*, 131–136.

Usoskin, D., Furlan, A., Islam, S., Abdo, H., Lönnerberg, P., Lou, D., Hjerling-Leffler, J., Haeggström, J., Kharchenko, O., Kharchenko, P. V, et al. (2014). Unbiased classification of sensory neuron types by large-scale single-cell RNA sequencing. Nat. Neurosci. *18*, 145–153.

Valentine, J.W. (2002). Cell-type number and complexity. Encycl. Evol. *1*, 144–146.

Vallejos, C.A., Marioni, J.C., and Richardson, S. (2015). BASiCS: Bayesian Analysis of Single-Cell Sequencing Data. PLoS Comput. Biol. *11*, 1–18.

Vallejos, C.A., Richardson, S., and Marioni, J.C. (2016). Beyond comparisons of means: Understanding changes in gene expression at the single-cell level. Genome Biol. *17*, 1–14.

Vallejos, C.A., Risso, D., Scialdone, A., Dudoit, S., and Marioni, J.C. (2017). Normalizing single-cell RNA sequencing data: Challenges and opportunities. Nat. Methods *14*, 565–571.

Vanderhaeghen, P., and Cheng, H.J. (2010). Guidance molecules in axon pruning and cell death. Cold Spring Harb. Perspect. Biol. *2*, 1–18.

Varley, J.E., and Maxwell, G.D. (1996). BMP-2 and BMP-4, but not BMP-6, increase the number of adrenergic cells which develop in quail trunk neural crest cultures. Exp. Neurol. *140*, 84–94.

Vickaryous, M.K., and Hall, B.K. (2006). Human cell type diversity, evolution, development, and classification with special reference to cells derived from the neural crest. Biol. Rev. Camb. Philos. Soc. *81*, 425–455.

Waddington, C.H. (1957). The strategy of the genes: A Discussion of Some Aspects of Theoretical Biology.

Wagner, A., Regev, A., and Yosef, N. (2016). Revealing the vectors of cellular identity with single-cell genomics. Nat. Biotechnol. *34*, 1145–1160.

Wagner, F., Yan, Y., and Yanai, I. (2018). K-nearest neighbor smoothing for high-throughput single-cell RNA-Seq data. bioRxiv 217737.

Wang, B., Zhu, J., Pierson, E., and Batzoglou, S. (2016). Visualization and analysis of single-cell RNA-seq data by kernel- based similarity learning. Bioarxiv *2*.

Wang, Y., Waters, J., Leung, M.L., Unruh, A., Roh, W., Shi, X., Chen, K., Scheet, P., Vattathil, S., Liang, H., et al. (2014). Clonal evolution in breast cancer revealed by single nucleus genome sequencing. Nature *512*, 155–160.

Weinreb, C., Wolock, S., Tusi, B.K., Socolovsky, M., and Klein, A.M. (2018). Fundamental limits on dynamic inference from single-cell snapshots. Proc. Natl. Acad. Sci. 201714723.

Wolf, F.A., Hamey, F., Plass, M., Solana, J., Dahlin, J.S., Gottgens, B., Rajewsky, N., Simon, L., and Theis, F.J. (2017). Graph abstraction reconciles clustering with trajectory inference through a topology preserving map of single cells. Doi.Org 208819.

Wolf, F.A., Angerer, P., and Theis, F.J. (2018). SCANPY: Large-scale single-cell gene expression data analysis. Genome Biol. *19*, 1–5.

Wu, C. -t., and Morris, J.R. (2001). Genes, Genetics, and Epigenetics: A Correspondence. Science (80-.). *293*, 1103–1105.

Yu, J., Xiao, J., Ren, X., and Xie, X.S. (2006). Probing Gene Expression in Live Single Escherichia coli Cells – One Molecule at a Time. Science (80-.). *311*, 1600–1603.

Zeisel, A., Munoz-Manchado, A.B., Codeluppi, S., Lonnerberg, P., La Manno, G., Jureus, A., Marques, S., Munguba, H., He, L., Betsholtz, C., et al. (2015). Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. Science (80-.). *347*, 1138–1142.

Zeisel, A., Hochgerner, H., Lonnerberg, P., Johnsson, A., Memic, F., van der Zwan, J., Haring, M., Braun, E., Borm, L., La Manno, G., et al. (2018). Molecular architecture of the mouse nervous system. bioRxiv.

Zheng, G.X.Y., Terry, J.M., Belgrader, P., Ryvkin, P., Bent, Z.W., Wilson, R., Ziraldo, S.B., Wheeler, T.D., McDermott, G.P., Zhu, J., et al. (2017). Massively parallel digital transcriptional profiling of single cells. Nat. Commun. *8*, 14049.

Ziegenhain, C., Vieth, B., Parekh, S., Reinius, B., Guillaumet-Adkins, A., Smets, M., Leonhardt, H., Heyn, H., Hellmann, I., and Enard, W. (2017). Comparative Analysis of Single-Cell RNA Sequencing Methods. Mol. Cell *65*, 631–643.e4.

žurauskiene, J., and Yau, C. (2016). pcaReduce: Hierarchical clustering of single cell transcriptional profiles. BMC Bioinformatics *17*, 1–11.