Computational and Comparative Analyses of Inhibitory Interneuron Diversity

vorgelegt von
M. Sc.
Joram Manasse Keijser

an der Fakultät IV - Elektrotechnik und Informatik
der Technischen Universität Berlin
zur Erlangung des akademischen Grades

Doktor der Naturwissenschaften
-Dr. rer. nat.-

genehmigte Dissertation

Promotionsausschuss:
Vorsitzender: Prof. Klaus Obermayer
Gutachter: Prof. Henning Sprekeler
Gutachterin: Prof. Susanne Schreiber
Gutachterin: Prof. Liset M de la Prida
Tag der wissenschaftlichen Aussprache: 11. März 2024

Berlin 2024

Zusammenfassung

Das Gehirn ist ein komplexes Netzwerk aus vielen Neuronen. Diese Neuronen können aufgrund ihrer vielfältigen molekularen, morphologischen und elektrophysiologischen Eigenschaften in Hunderte von Zelltypen eingeteilt werden. Diese verschiedenen Zelltypen spielen bei neuronalen Berechnungen eine bedeutende Rolle. Um die Funktionsweise des Gehirns zu verstehen, ist es daher wichtig, die verschieden Zelltypen und deren Zusammenwirken zu erforschen.

Eine besonders vielfältige Gruppe von Zellen ist die der inhibitorischen Interneurone, benannt nach ihrer Wirkung auf die lokale neuronale Aktivität. Diese Interneurone inhibieren andere Neurone, mit denen sie verbunden sind und stabilisieren damit die neuronale Aktivität in einem Netzwerk. Warum dafür jedoch viele verschiedene inhibitorische Zelltypen erforderlich sind, ist unklar.

In unserer ersten Studie haben wir die Hypothese aufgestellt, dass verschiedene inhibitorische Zelltypen benötigt werden, um die neuronale Aktivität in den verschiedenen Abschnitten exzitatorischer Pyramidenzellen (Soma und Dendriten) zu stabilisieren. Um diese Hypothese zu testen, optimierten wir die Eigenschaften von Modell-Interneurone, sodass die exzitatorischen (E) und inhibitorischen (I) Eingangsströme an den verschiedenen Abschnitten einer exzitatorischen Pyramidalzelle im Gleichgewicht sind. Nach der Optimierung ließen sich die Neurone weitgehend in zwei Klassen einteilen, die den biologischen Interneuron-Typen, den sogenannten PV+- und SST+-Neuronen, ähneln. Diese Zelltypen könnten daher für die Regulierung eines E/I Gleichgewichtes in verschiedenen Abschnitten eines Neurons erforderlich sein.

Eine naheliegende Schlussfolgerung ist daher, dass sich die Zelltypen PV+ und SST+ zu diesem Zweck evolutive entwickelt haben könnten. In unserer zweiten Studie haben wir diese Idee anhand von Zelltypen verschiedener Tierarten getestet. Überraschenderweise fanden wir Hinweise darauf, dass Interneurone die Fähigkeit zur abschnittsspezifischen Inhibition erworben haben, bevor es einen Selektionsdruck für diese Funktion gab. Die Eigenschaften eines Zelltyps könnten also im Laufe der Evolutionsgeschichte zusammengewürfelt worden sein, anstatt dass sie für einen Zweck optimiert wurden.

Des Weiteren untersuchten wir, wie sich die Aktivität verschiedener Interneuron-Typen mit dem Verhalten eines Tieres ändert. Es wurde für Mäuse gezeigt, dass einige Interneurone ihre Aktivität erhöhen, wenn das Tier aktiv ist, während andere inhibitorische Zelltypen ihre Aktivität reduzieren. Um zu prüfen, ob dies auch bei anderen Tierarten der Fall ist, müssten sowohl die Aktivität als auch die Identität der einzelnen Zellen gemessen werden. Leider ist dies derzeit nicht möglich, insbesondere nicht beim Menschen. Daher haben wir in den leichter verfügbaren Daten zur Genexpression nach Mustern gesucht, die mit solchen Bewegungskorreliereten Neuronenaktivitäten in Verbindung stehen. Wir entdeckten, dass diese Muster beim Menschen recht ähnlich zu denen der Mäuse sind, sich aber signifikant bei

Schildkröten- und Zebrafinken-Interneuronen unterscheiden. Anhand eines mathematischen Modells machen wir Vorhersagen, welche dieser Genexpressionsunterschiede sich in funktionelle Unterschieden niederschlagen können.

Zusammenfassend bietet diese Dissertation eine nuancierte Perspektive auf die Interneuronenvielfalt. Verschiedene Zelltypen haben unterschiedliche Funktionen, aber diese Funktionen sind nicht unbedingt der Grund für ihre Existenz. Interneurone verschiedener Arten sind evolutionär verwandt, haben aber natürlich auch ihre eigenen Spezialisierungen entwickelt.

Abstract

The brain is a complex network of many brain cells, in particular neurons. These neurons are very diverse and can be categorized into hundreds of cell types based on their molecular, morphological, and electrophysiological properties. Because different cell types play specialized roles in neural computations, understanding the brain requires understanding its cell types and how they interact.

One particularly diverse group of cells is that of inhibitory interneurons, named after their suppressive effect on local neural activity. Interneurons are important to stabilize neural activity, but why this requires many different cell types is unclear.

In our first study, we hypothesized that different cell types are needed to stabilize neural activity in the various compartments (soma and dendrites) of excitatory pyramidal cells. To test this hypothesis, we optimized the properties of model interneurons to balance excitation (E) and inhibition (I) in both pyramidal compartments. After the optimization, neurons largely fell into two classes, resembling biological interneuron types known as PV and SST-positive neurons. These cell types may therefore be required for balancing excitation and inhibition in different neuronal compartments.

A natural conclusion, then, is that PV+ and SST+ cell types might have developed or evolved for this purpose. In our second study, we tested this idea by comparing cell types across species. Surprisingly, we found evidence that interneurons acquired the ability of compartment-specific inhibition before there was selective pressure for this function. The properties of a cell type may, therefore, have been cobbled together during its evolutionary history rather than being optimized for its current purpose.

Finally, we investigated the behavioural response of different interneuron types. When a mouse becomes active and aroused, certain interneuron types also become more active, but others become less active. Testing whether this is true in other species would require measuring both the activity and identity of individual cells. Unfortunately, this is currently unfeasible, especially in humans. We therefore searched more readily available gene expression data for patterns associated with arousal responses. We discovered that these patterns are quite similar in humans but different in turtle and zebra finch interneurons. A mathematical model predicts which of those gene expression differences may translate into functional differences and which may not.

In summary, this thesis provides a nuanced perspective on interneuron diversity. Different cell types have distinct functions, but these functions may not be the reason that the cell types exist. Interneurons of different species are evolutionarily related but have, of course, also developed their own specializations.

... none of the materials at the tinkerer's disposal has a precise and definite function. Each can be used in a number of different ways. In contrast with the engineer's tools, those of the tinkerer cannot be defined by a project. What these objects have in common is "it might well be of some use." For what? That depends on the opportunities.

—Francois Jacob, *Evolution and Tinkering*

Acknowledgements

On an otherwise unremarkable day in October 2017, I met with Henning Sprekeler to discuss a rotation in his lab. By the end of the meeting, Henning's bombardment of incisive questions had made it abundantly clear how much I could learn from him if only he'd take me on as his student. I am glad that he did. Six years later, I am still as impressed by Henning's keen intellect and desire to understand how things work, but I've equally come to appreciate his empathy and kindness. I am particularly grateful for the freedom he gave me to explore my interests inside and outside his lab.

I am also grateful for the support from other members of the lab. Loreen Hertäg has become a wonderful mentor whose guidance and encouragement have been invaluable; I can only wish for a fraction of her conscientiousness and planning skills. Working with Laura Naumann was a highlight of my PhD, and I hope it has taught me some of her goal-directedness and good spirit. Filip Vercruysse has positively influenced me through his love of science and running—Filip's remarkable optimism and perseverance always brighten my mood. Robert Lange must be the most driven yet warmest person I know, and he is a true inspiration. Soon, Robert will be wildly successful, and I will tell people I always knew he would be. Denis Alevi's presence and automation skills have greatly improved the fun and efficiency of my teaching efforts. I would also like to thank Mark Boon, Cathrin Bunkelmann, Felix Lundt, Owen Mackwood, Yigit Oksuz, Friedrich Schuessler and Soledad Traverso for contributing to the positive atmosphere that makes the Sprekeler Lab such a great place. Special thanks to Filip, Friedrich, Laura, and Loreen for their feedback on this thesis.

Outside the Sprekeler Lab, I thank Anna Schroeder and Johannes Letzkus for their collaboration. I am also grateful to Simon Butt for kindly sharing his insights on neural development. I thank Liset de la Prida and Susanne Schreiber for their time and effort in reviewing this thesis and Klaus Obermayer for chairing the thesis defence.

On a personal level, I am indebted to Elianne for convincing me to apply for the ECN fellowship and for her selfless support when I was lucky enough to receive it. Thanks to Polina for hosting me during my first weeks in Berlin. I am grateful to Maxime for keeping in contact and being the sincere friend that she is. I appreciate Adrian for our library sessions and the Berlin library employees for keeping things open on the weekend. I want to acknowledge my brothers, Amos and Ruben, whose passions for literature, sports, and perhaps even social issues have been a great inspiration. Thanks to my parents, Job and Karin, for nurturing my curiosity and for their continuing love. I hope they can live with all the tinkering in this thesis. Finally, I thank Alexandra for supporting me in good and bad times, infinitely expanding and improving my life outside work, and turning Berlin into a home.

Table of Contents

List of Figures

List of Tables

1

Scientific background

As soon as scientists found a way to look at individual neurons, they became fascinated by their diversity. Peering through a microscope, Santiago Ramón y Cajal created astonishingly detailed drawings of neurons and their various, until then largely unnamed and even unknown, parts (Fig. 1.1). Once an aspiring painter, Cajal used his drawings to show that neurons come in many shapes. He also inferred that within each neuron, information flows through specialized input and output processes.

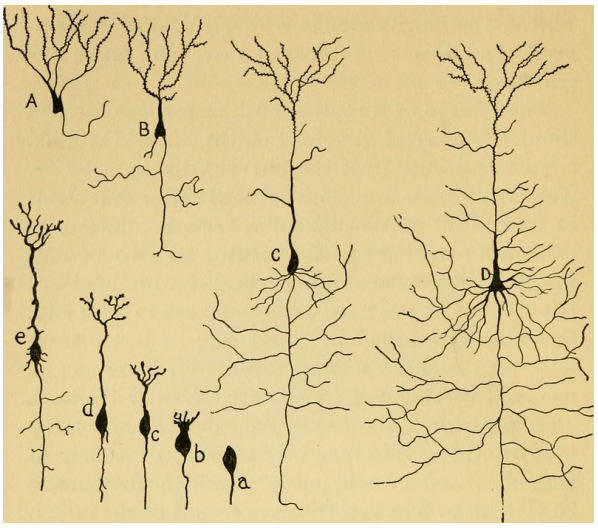

Figure 1.1: Ramón y Cajal: Evolution and development of the nervous system. **A-D**, neurons from different species; **a-e**, neurons across stages of development. Figure from `commons.wikimedia.org`, originally published in Thomson (1925) and released into the public domain.

This diversity raises fundamental questions: what functions does it serve? How do a neuron's input and output processes influence its information processing? And how did all this complexity evolve and develop? Answers to these questions have begun to emerge in the century since Cajal. We now know, for example, that neurons communicate through excitatory or inhibitory neurotransmitters, glutamate and GABA (γ-aminobutyric acid), respectively.

Glutamatergic projection neurons often establish long-distance connections, while GABAergic interneurons form local networks.

For practical reasons, neuroscience has largely concentrated on excitatory projection neurons rather than inhibitory interneurons. Not only are interneurons small and rare, but they are also incredibly diverse, and this hinders their experimental investigation. However, recent decades have witnessed a shift as technological advances allow neuroscientists to investigate even the rarest interneurons. New methods for profiling gene expression at the single-cell level have revealed an even greater neuronal diversity than previously appreciated. The recording and manipulation of specific interneuron types' activities reveal distinctive roles in sculpting cortical information flow. Simultaneously developed computational methods enable the analysis of these complex data.

While incredible progress has been made since the time of Cajal, numerous questions persist concerning neuronal diversity, its functions, and its origins. This thesis addresses some of these questions through theoretical modelling and data analysis. First, we show how interneurons can control information flow through a division of labour that mirrors the complex physiology of pyramidal cells. Next, we ask how this intricate interplay of excitation and inhibition evolved by comparing neuronal properties across species. Finally, we use transcriptomic analyses to understand how diverse interneuron types collectively regulate cortical information flow.

Two introductory chapters will review the scientific and methodological background while highlighting the key themes running through the thesis.

1.1 Excitatory pyramidal cells are highly complex

Early neuroscientists were particularly intrigued by the elaborate branching processes of neurons. Some of Cajal's most iconic illustrations feature pyramidal cells, showcasing numerous basal dendrites radiating from the cell body and a single apical dendrite extending towards the upper cortical layers. Pyramidal neurons intrigued Cajal to the extent that he referred to them as "psychic cells", the privileged substrate of mental functions (Ramón y Cajal, 1894).

Subsequent electrophysiological and imaging experiments have confirmed that their dendrites indeed endow pyramidal neurons with unique, although not necessarily psychic, functions (London and Häusser, 2005; Payeur et al., 2019). Instead of being passive relays for information transmission to the cell body, dendrites contain voltage-gated ion channels that amplify local inputs (Yuste et al., 1994; Stuart et al., 1997; Williams and Stuart, 2002). This allows groups of dendrites to act as functional "subunits" (Polsky et al., 2004), potentially granting single pyramidal neurons the computational capabilities of an entire artificial network (Poirazi et al., 2003; Beniaguev et al., 2021). While the extent to which dendrites operate independently is unresolved (Beaulieu-Laroche et al., 2019; Francioni et al., 2019), it is clear that their complexity expands the computational repertoire of pyramidal neurons.

The following example will play an important role in Chapters 4 and 5 of this thesis. Under typical conditions, somatic spikes are significantly attenuated when propagating into the dendritic tree. But the same somatic stimulation creates a sustained dendritic calcium spike when paired with dendritic stimulation (Larkum et al. (1999); Fig. 1.2). This "plateau

potential" propagates back to the soma (Stuart and Sakmann, 1994), triggering a burst of additional somatic spikes (Larkum et al., 1999). Single spikes and bursts could therefore form two qualitatively different signals: Whereas a single spike indicates somatic activity, a burst indicates simultaneous somatic and dendritic activity (Larkum et al., 2001; Shai et al., 2015).

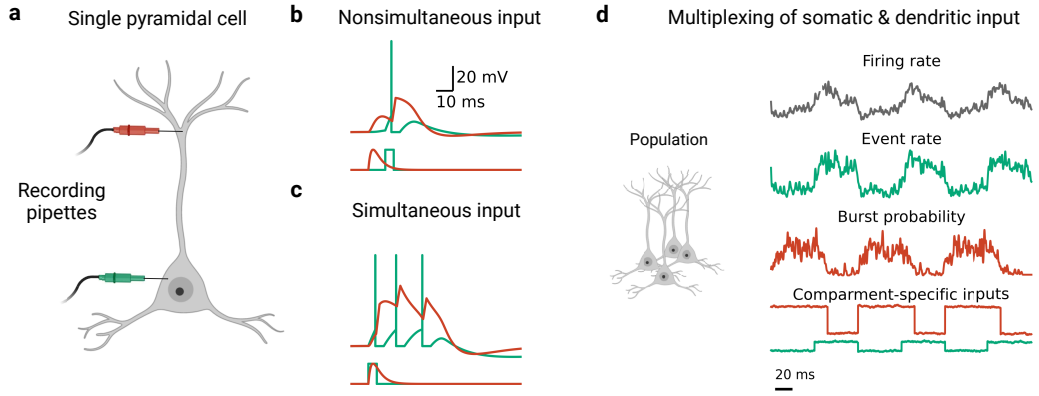

Figure 1.2: Pyramidal physiology and burst multiplexing. (a) Simultaneous recordings in the soma and apical dendrite of a layer 5 pyramidal cell inspired by those from Larkum et al. (1999). Somatic stimulation will lead to a single spike **(b)**, but coincident somatic and dendritic input cause a dendritic plateau potential and a burst of multiple somatic spikes **(c)**. **(d)** Across a population of cells, the event rate and burst probability independently reflect somatic and dendritic inputs, respectively. Simulations replicated from Naud and Sprekeler (2018); icons from Biorender.

In a recent theoretical study, Naud and Sprekeler (2018) proposed how pyramidal cells might simultaneously communicate somatic and dendritic information sources. In their model, a cell's dendritic input is reflected in the relative number of bursts compared to single spikes (burst probability). A cell's somatic input, on the other hand, is reflected in the rate of both bursts and single spikes. As an individual pyramidal cell either bursts or does not burst, the burst probability is transmitted through a population of many cells (Fig. 1.2d). Downstream circuits could tell apart spikes from bursts using a combination of short-term plasticity and disynaptic inhibition (Naud and Sprekeler, 2018). As we show in Chapter 4, this could allow interneurons to form compartment-specific feedback loops, routing, e.g., dendritic activity back to the dendrite.

Multiplexed transmission of somatic and dendritic inputs would not be particularly useful unless these inputs were actually distinct, which they appear to be. Whereas perisomatic compartments receive feedforward inputs, apical dendrites receive feedback information (Petreanu et al. (2009)—although this separation is less clear-cut than is sometimes assumed; Young et al. (2021) and Ledderose et al. (2023)). In the sensory cortex, for example, feedforward and feedback inputs might correspond to sensory inputs and predictions of those inputs, respectively (Rao and Ballard, 1999; Keller and Mrsic-Flogel, 2018; Hertäg and Sprekeler, 2020). The unique input organization and physiology of pyramidal cells could enable them to perform the crucial task of coupling feedforward and feedback information (Xu et al., 2012; Larkum, 2013). Segregated cellular compartments might also implement the backpropagation algorithm without the need for separate forward and backward passes (Körding and König (2000), Sacramento et al. (2018), Payeur et al. (2021), and Richards et al. (2019)).

1.2 Inhibitory interneurons are highly diverse

Consistent with their unique properties, distinct pyramidal compartments are innervated by different types of interneurons (Markram et al., 2004; Tremblay et al., 2016; Fishell and Kepecs, 2020). A pyramidal cell's soma and perisomatic dendrites are inhibited by cells that express the parvalbumin (PV) gene (Hu et al., 2014). A pyramidal cell's distal dendrites, on the other hand, are the target of cells that express the somatostatin (SST) gene (Liguz-Lecznar et al., 2016; Wang et al., 2004). SST-positive cells exert a powerful influence on dendritic activity: they can block the initiation of calcium potentials (Larkum et al., 1999; Silberberg and Markram, 2007) and bursting (Lovett-Barron et al., 2012; Gentet et al., 2012) in a localized manner (Chiu et al., 2013). PV+ and SST+ interneurons also have specific connectivity patterns with other interneuron types (Pfeffer et al., 2013; Jiang et al., 2015; Campagnola et al., 2022). Whereas PVs inhibit pyramidal cells and each other, SSTs inhibit other interneuron types but not themselves. PV+ cells receive feedforward excitation (Yu et al., 2019; Naskar et al., 2021), but SST+ cells are rather recruited by local inputs (Gentet et al., 2012; Yu et al., 2019) and, depending on the area, top-down inputs (Shen et al., 2022).

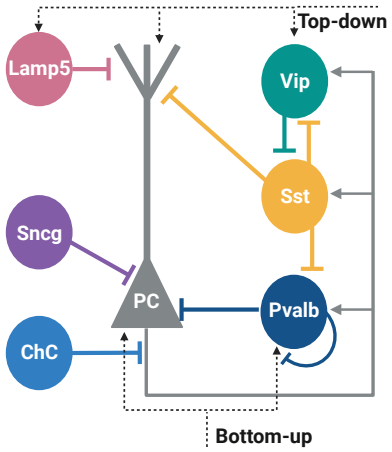

Figure 1.3: Cortical microcircuit of a pyramidal cell (PC) and multiple interneuron types. The circuit receives sensory bottom-up inputs and associative top-down inputs. Pvalb, Sst, Vip, Sncg, and Lamp5 are marker genes, ChC is a chandelier cell. A few connections are left out for visual clarity.

Although PVs and SSTs receive recurrent excitation from a potentially overlapping set of pyramidal cells, the dynamics of these signals are markedly different (Reyes et al., 1998; Silberberg and Markram, 2007). Synapses arriving at a PV+ cell are short-term depressing, i.e. consecutive spikes have a progressively weaker effect. Synapses onto SST+ cells, on the other hand, are short-term facilitating. SST+ cells will therefore be preferentially recruited by bursts of multiple spikes.

The majority of this thesis will focus on PV+ and SST+ cells. This focus is not entirely unjustified since these are some of the most prevalent cortical interneuron types (Rudy et al., 2011). But they are by no means the only ones. For one thing, the family of PV+ and SST+ cells contains many different subtypes, some of which do not target the soma or apical dendrites. For example, PV+ Chandelier cells are fast-spiking but control the pyramidal initial axonal segment (Somogyi, 1977; Inan and Anderson, 2014). SST+ basket cells lack dendritic

projections, are fast spiking and inhibit excitatory stellate cells in layer 4 (Nigro et al., 2018; Scala et al., 2019; Naka et al., 2019). Since marker genes such as PV+ and SST+ do not label homogenous cell types, a complete characterisation of interneuron diversity requires the measurement of many genes.

Interneuron diversity also comprises entirely different cell classes beyond PV+ and SST+ cells (Fig. 1.3). VIP-expressing neurons specialise in disinhibition by targeting other interneurons (Pi et al., 2013; Lee et al., 2013; Pfeffer et al., 2013), in particular SST+ cells. The class of LAMP5-expressing cells is mainly found in upper cortical layers (Tasic et al., 2018). A subset of LAMP5+/NDNF+ cells has recently received attention as a second source of dendritic inhibition (Abs et al., 2018; Malina et al., 2021; Hartung et al., 2023). NDNFs are inhibited by SSTs (Abs et al., 2018) and inhibit PVs (Malina et al., 2021). Finally, SNCG-expressing neurons are also found in upper cortical layers. Like PVs, certain SNCGs inhibit the perisomatic region of pyramidal cells, although with different temporal dynamics (Dudok et al., 2021) and with a bias for different pyramidal types (Jézéquel et al., 2023). In mouse visual cortex, SNCG+ cells are inhibited by visual stimulation but excited by locomotion (Bugeon et al., 2022).

The cortex therefore contains many different types of interneurons characterized by combinations of different properties. Which purpose could all this diversity serve?

1.3 Excitation/inhibition balance: a key function of interneurons

One of inhibition's key functions is to regulate neural activity by balancing excitation (Isaacson and Scanziani (2011); Fig. 1.4). Balanced, i.e. temporally or spatially coinciding, excitation and inhibition have indeed been widely observed (Wehr and Zador, 2003; Haider et al., 2006; Okun and Lampl, 2008; Bhatia et al., 2019). Balanced excitation and inhibition can create the asynchronous and irregular activity often observed in cortical networks (Van Vreeswijk and Sompolinsky, 1996; Brunel, 2000; Renart et al., 2010) and enable rich and rapid dynamics through selective amplification (Murphy and Miller, 2009; Hennequin et al., 2014).

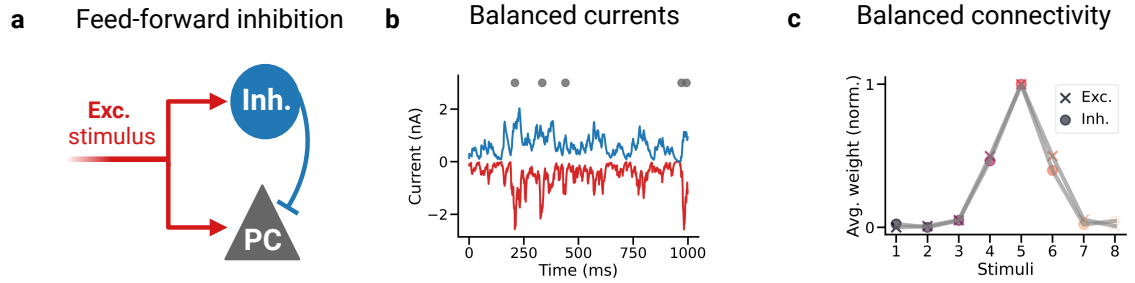

Figure 1.4: Excitation/inhibition balance in a small network. **(a)** Feed-forward inhibition: a pyramidal cell (grey) receives inhibition from interneurons (Inh.) that are driven by the same excitatory stimulus. **(b)** Inhibitory currents track excitatory currents on short-timescales. Dots indicate spike times. **(c)** Excitation and inhibition synaptic strengths are balanced across multiple stimuli. Simulations after Vogels et al. (2011).

In cortical networks, PV+ cells are considered the primary stabilizing force. PVs provide fast feedforward inhibition (Estebanez et al., 2017; Yu et al., 2019; Nashef et al., 2022),

track local network activity (Dipoppa et al., 2018), and match inhibitory to excitatory inputs across pyramidal cells (Xue et al., 2014). SST+ cells complement PVs by providing lateral inhibition (Silberberg and Markram, 2007; Stefanelli et al., 2016) and by specializing in dendritic control (Liguz-Lecznar et al., 2016; Yavorska and Wehr, 2016). SST+ cells do this in a spatially precise manner (Chiu et al., 2013), with different subtypes inhibiting specific subcellular compartments (Wu et al., 2023). VIP+ cells play a fundamentally different role from PV+ and SST+ cells: they inhibit SST+ cells and can thus disinhibit pyramidal dendrites (Pi et al., 2013; Karnani et al., 2016a; Naskar et al., 2021) to allow signal propagation (Yang et al., 2016; Wilmes and Clopath, 2019).

Although the term excitation-inhibition balance might evoke a fixed and rigid control system, the opposite is true: inhibition dynamically adapts to behavioural states (Batista-Brito et al., 2018; Cardin, 2019). During quiet wakefulness, cortical circuits exhibit slow, correlated oscillations, with PV+ cells being highly active. In contrast, periods of high arousal are characterised by desynchronized activity, and cell types such as SST+ and VIP+ cells become dominant. This state-dependent activity is partially mediated by neuromodulators such as acetylcholine (Metherate et al., 1992; Poorthuis et al., 2014; McCormick et al., 2020). Acetylcholine hyperpolarizes PVs (Xiang et al., 1998; Alitto and Dan, 2013) and depolarizes SSTs (Xiang et al., 1998; Kawaguchi, 1997; Chen et al., 2015b). Since acetylcholine simultaneously activates VIPs (Kawaguchi, 1997; Chen et al., 2015b; Gasselin et al., 2021), its ultimate impact on SSTs may depend on the strength of inhibition from VIP+ cells on one hand, and receptor expression on the other (Muñoz et al., 2017). The cholinergic shifting of inhibition from soma to dendrites could enhance sensory "bottom-up" information over associative "top-down" modulation (Hasselmo, 1995; Bugeon et al., 2022). We will investigate this cholinergic reweighing of inputs in Chapter 6.

It is currently unclear to what extent excitation-inhibition balance is a useful framework for understanding circuit-level functions of other interneuron classes. NDNF+ neurons broadly inhibit dendrites (and PV+ cells) on a slow timescale (Abs et al., 2018; Hartung et al., 2023). SNCG+ cells inhibit PCs but are anticorrelated with their activity in hippocampus (Dudok et al., 2021) and visual cortex (Bugeon et al., 2022). PV+ chandelier cells may have a weak inhibitory (Seignette et al., 2023) or even excitatory (Szabadics et al., 2006) effect. Distinct interneuron types therefore control cortical information flow in unique ways without necessarily relying on a precise spatiotemporal relationship with excitatory activity.

2

Methodological background

Individual neurons are highly complex and diverse, yet their assembly into networks ultimately gives the nervous system its computational power. Exactly how this happens is a key question facing neuroscientists today. The answer cannot be found simply by "reasoning through" the problem: feedback loops and nonlinear dynamics make the system too complex. Instead, understanding the brain requires quantitative models.

2.1 Network models describe neuronal interactions

The evolution of neural activity is naturally modelled as a dynamical system. In this approach, mathematical equations describe how neural activity changes depending on itself and external inputs. Here, we will often be interested in the distinction between activity in the form of single spikes and bursts of multiple spikes. A simple model that allows for this distinction is the leaky integrate and fire (LIF) neuron (Brunel and Van Rossum, 2007; Gerstner et al., 2014). This model integrates (or sums) its inputs and fires an action potential upon crossing a threshold from below:

$$\tau \frac{dV_i}{dt} = -V_i(t) + \sum_j w_{ij}^{rec} S_j(t) + \sum_j w_{ij}^{in} I_j(t) \tag{2.1}$$

$$V_i(t) \geq 1 \implies V_i(t) := 0 \text{ and } S_i(t) := \delta(t). \tag{2.2}$$

Here, V_i is neuron i's membrane potential, which exponentially relaxes to 0 without input. If the voltage crosses the threshold value of 1, it is reset, and the neuron fires a spike, recorded as a delta function $\delta(t)$. After a spike, the voltage is reset to 0, creating a discontinuous trajectory. In practice, network models are simulated using a discretisation scheme like Euler's method. This discretizes time into small bins of length Δt:

$$V_i(t + \Delta t) = V_i(t) + \frac{\Delta t}{\tau} \times \left(-V_i(t) + \sum_j w_{ij}^{rec} S_j(t) + \sum_j w_{ij}^{in} I_j(t) \right). \tag{2.3}$$

Extending this model with dendritic nonlinearities can be done using a second, dendritic equation (Naud et al., 2014). Other extensions include a slow-timescale variable for spike-frequency adaptation and the explicit modelling of current or conductance-based synapses (Gerstner et al., 2014; Vogels and Abbott, 2005).

Finally, we want to model a key feature of interneuron types: short-term plasticity. We will use the Tsodyks-Markram model (Tsodyks et al., 1998; Costa et al., 2013), which models short-term plasticity as the interplay of a synaptic release probability u and a number of released vesicles r. The product of release probability and released vesicles determines the time-varying weight scale $\mu(t) = u(t)r(t)$. This changes Eq. (2.1) to:

$$\tau \frac{dV_i}{dt} = -V_i(t) + \sum_j w_{ij}^{rec} \cdot \mu_{ij}(t) \cdot S_j(t) + \sum_j w_{ij}^{in} I_j(t). \tag{2.4}$$

The u and r equations are explained in the methods section of Chapter 4. This model can accommodate both facilitation and depression, depending on the short-term plasticity parameters.

As is clear from Eq. 2.4, including biological features can easily lead to complicated models. We will therefore need a clever way to set the network's parameters, in particular its connectivity.

2.1.1 Machine learning can model complex neuronal functions

Classically, the connectivity of neural network models is hard-wired or learned using local plasticity rules. Although this approach has led to important insights (see, e.g., Brunel (2000), Wong and Wang (2006), and Song et al. (2000)), it struggles to create models that perform complex tasks.

Machine learning provides a powerful way to solve a computational problem while skipping, at least initially, the need to intuit a solution. Instead, the approach quantifies functional success with a reward or cost function and optimises this objective by systematically searching over network parameters. The modeller thus specifies *what* the model should do, but not *how* (Sussillo, 2014; Barak, 2017). The basic idea is not new: biologists have long used optimisation theory to understand the costs and benefits of particular phenotypes (Smith, 1990; Alexander, 1996; Bialek, 2012; Alon, 2019). What is new, however, is the increased capabilities of machine learning methods, in particular deep neural networks.

Deep neural networks are typically trained using gradient-based optimization (LeCun et al., 2015). In this approach, gradients are calculated using the backpropagation algorithm (Rumelhart et al., 1986; Werbos, 1990), which uses the chain rule to compute each parameter's contribution to the objective function. This so-called backward pass retraces the individual computations that generated the model's output in the forward pass (Fig. 2.1). Crucially, automatic differentiation software eliminates the need to implement backpropagation by hand (Paszke et al., 2017; Baydin et al., 2018; Bradbury et al., 2018). Instead, gradients are obtained "for free" after implementing the forward pass. This allows the optimization of any parameter, including synaptic weights, membrane time constants, and initial release probabilities. Computational neuroscientists can thus model the neural implementation of a specific computation in a few conceptually straightforward steps:

1. Model the task, for example, by collecting or simulating the input stimuli and corresponding behavioural responses. Well-known tasks include image categorisation and decision-making.

2. Decide on a model class, such as a fully connected RNN, and obtain its response to the input stimuli.

3. Quantify task performance using a mathematical objective function such as the mean squared difference between the correct and actual network outputs.

4. Optimize the network model using automatic differentiation.

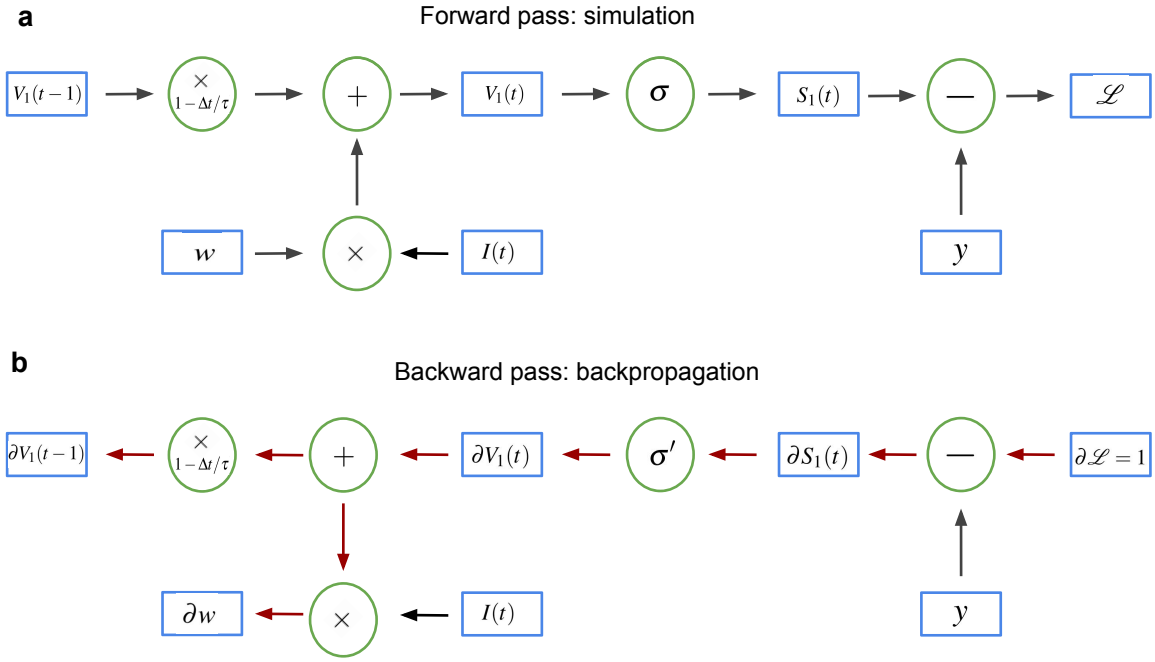

Figure 2.1: Computing gradients using backpropagation. (a) A discretized version of the continuous system Eq. (2.1) is simulated during the forward pass. The final step consists of computing a scalar loss $\mathcal{L}$ that compares the network output with a target y. Only a single neuron is shown for simplicity. **(b)** The backward pass traces the same computational graph in reverse order, starting from the loss $\mathcal{L}$. Each step now computes the derivative or gradient of the forward computation and combines this with the other derivatives using the chain rule. ∂w is a shorthand for $\partial L/\partial w$, the derivative or gradient of the loss function $\mathcal{L}$ with respect to the weight w.

This approach has been used to model decision making (Mante et al., 2013; Molano-Mazón et al., 2023), vision (Zipser and Andersen, 1988; Yamins et al., 2014), motor control (Sussillo et al., 2015; Feulner and Clopath, 2021), timing (Wang et al., 2018; Beiran et al., 2023), navigation (Cueva and Wei, 2018; Banino et al., 2018), working memory (Barak et al., 2013; Orhan and Ma, 2019), and many other domains (Richards et al., 2019; Yang and Wang, 2020). Machine learning is also used to fit network models to empirical activity patterns (Pandarinath et al., 2018; Perich et al., 2020), and to track animal behavior (Mathis et al., 2018; Pereira et al., 2020). In short, machine learning allows neuroscientists to model how the brain performs complex computations. Although we will later point out important limitations of an optimization-based approach to brain research, it is difficult to underestimate its value for contemporary neuroscience.

2.1.2 Optimising detailed models requires special care

Although machine learning allows the modelling of biological computations, this does not automatically result in a good model. After training, a deep network model must therefore be analysed and compared with neural (or behavioural) data. This comparison is often made at the abstract level of neural manifolds (Gallego et al., 2017; Vyas et al., 2020) and their geometry (Kriegeskorte et al., 2008; Chung and Abbott, 2021), although the activity of individual neurons can be predicted as well (Sussillo et al., 2015; Yamins et al., 2014).

Trained networks thus serve as an abstract model of neural function without much regard for the underlying structure. For example, RNNs often use gated architectures (Hochreiter and Schmidhuber, 1997; Cho et al., 2014) that are easier to train but less biologically realistic than fully connected alternatives. Moreover, a network's optimised connectivity matrix is rarely compared to biological observations, and neuronal parameters are typically not optimised (see Perez-Nieves et al. (2021) and Geadah et al. (2022) for exceptions). Therefore, deep learning has not enjoyed the same success in explaining brain structure as brain function. The limited success may partially stem from the lack of efficient algorithms for optimising detailed networks, especially those with spiking neurons. Gradient descent gradually changes parameters in the right direction and therefore requires a smooth relationship between parameters and performance. But for a spiking neuron, most small input changes will not change its output from silent to spiking or vice versa. Therefore, the derivative of the threshold spiking operation is zero (or not defined at the spiking threshold).

A recently proposed solution—surrogate gradient descent—replaces the non-differentiable spiking function with a smooth approximation (Zenke and Ganguli (2018), Neftci et al. (2019), and Bellec et al. (2020); Fig. 2.2). In the forward pass, the network is simulated as usual, but in the backward pass, the spiking function's derivative is replaced with that of a smooth approximation, such as a sigmoid. Although this approach results in a biased gradient estimate, at least the estimate is nonzero and, hopefully, aligned with a descent direction. It is currently an open question to what extent biased gradients hurt the performance of the surrogate gradient approach. It is also unclear whether gradient-based optimization is even preferable over gradient-free methods for detailed neuron models (Hazelden et al., 2023). Rigorous benchmarks (Zenke and Vogels, 2021; Cramer et al., 2020) and exact algorithms (Wunderlich and Pehle, 2021; Klos and Memmesheimer, 2023) may answer this question. Here, we will use surrogate gradient optimisation to explain the structure of interneuron circuits (Chapter 4).

2.2 Single-cell RNA-seq: genome-wide expression with cellular resolution

Traditionally, neuronal types have been defined based on a small number of properties, such as morphology or the expression of a single marker gene like PV or SST. However, any one gene is rarely expressed in only one cell type. Instead, cell types are characterised by the combinatorial expression of many genes. The definition of cell types thus requires high-dimensional and large-scale measurements. A good candidate method is single-cell RNA sequencing (scRNA-seq; Tang et al. (2009)), which measures how many copies of each gene are expressed by individual cells. ScRNA-seq's high throughput and resolution are revolutionising many parts of biology,

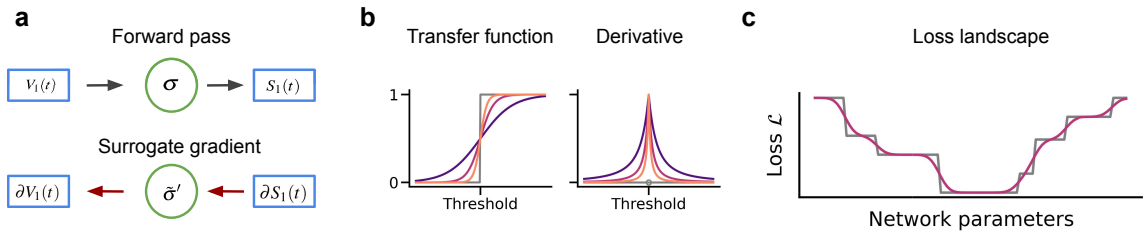

Figure 2.2: Surrogate gradient descent can optimise spiking networks. (a) The forward pass uses the "true" spiking function as threshold-nonlinearity σ. During the backward pass, the derivative of the spiking function is replaced by that of a smooth approximation $\tilde{\sigma}$. **(b)** Example of a widely-used surrogate gradient, parametrized by a slope that trades off bias with gradient flow. **(c)** By descending a smoothed loss landscape, surrogate gradient descent might avoid getting stuck.

including cellular neuroscience (Aldridge and Teichmann, 2020; Stark et al., 2019; Tasic, 2018). The resolution and throughput of scRNA-seq have allowed the detailed categorisation of cell types (Zeng, 2022). This has revealed a hierarchical organization (Tasic et al. (2018) and Yao et al. (2021); Fig. 2.3). The next sections will review the acquisition and analysis of scRNA-seq data.

2.2.1 ScRNA-seq data is generated in several steps

To collect scRNA-seq data, researchers must capture the tiny amount of RNA in individual cells and sequence it across many cells in parallel. The process involves several steps (Haque et al. (2017) and Hwang et al. (2018); Fig. 2.4). First, single cells are isolated from the tissue of interest, such as a particular brain area. Next, individual cells are separated to prevent the mixing of their RNA molecules. The cell membrane is lysed to release the RNA, and each molecule is assigned a unique barcode that enables its tracking through the sequencing process. Next, the RNA is transcribed into complementary DNA (cDNA) and amplified to generate enough material. This cDNA is equivalent to the RNA but can be sequenced using a massively parallel sequencing platform (Metzker, 2010; Goodwin et al., 2016). After sequencing has determined the DNA bases of each amplified fragment, the fragments are assigned to specific genes by comparison with a reference genome (Li and Homer, 2010; Alser et al., 2021). This determines the number of detected fragments from each gene, which is summarised in a cell-by-gene count matrix. Depending on the experimental setup, the data typically contains 1000 to 10,000 cells and ca. 20,000 genes.

The specific steps of a scRNA-seq experiment depend on the experimental protocol (Kolodziejczyk et al., 2015), which has important implications for downstream data analysis. One variation involves sequencing the RNA of single nuclei rather than whole cells (single nucleus (sn)RNA-seq). This method yields a lower RNA count (Bakken et al., 2018) but does not require fresh tissue since nuclei can be better preserved (Krishnaswami et al., 2016; Lacar et al., 2016). Experimental protocols also differ in the strategy for isolating single cells and in the portion of RNA that they analyse. Some protocols isolate cells into individual wells and capture the entire length of RNA transcripts (e.g., Picelli et al. (2014)). Others use a droplet-based method to capture one end of each RNA sequence (e.g., Macosko et al. (2015)).

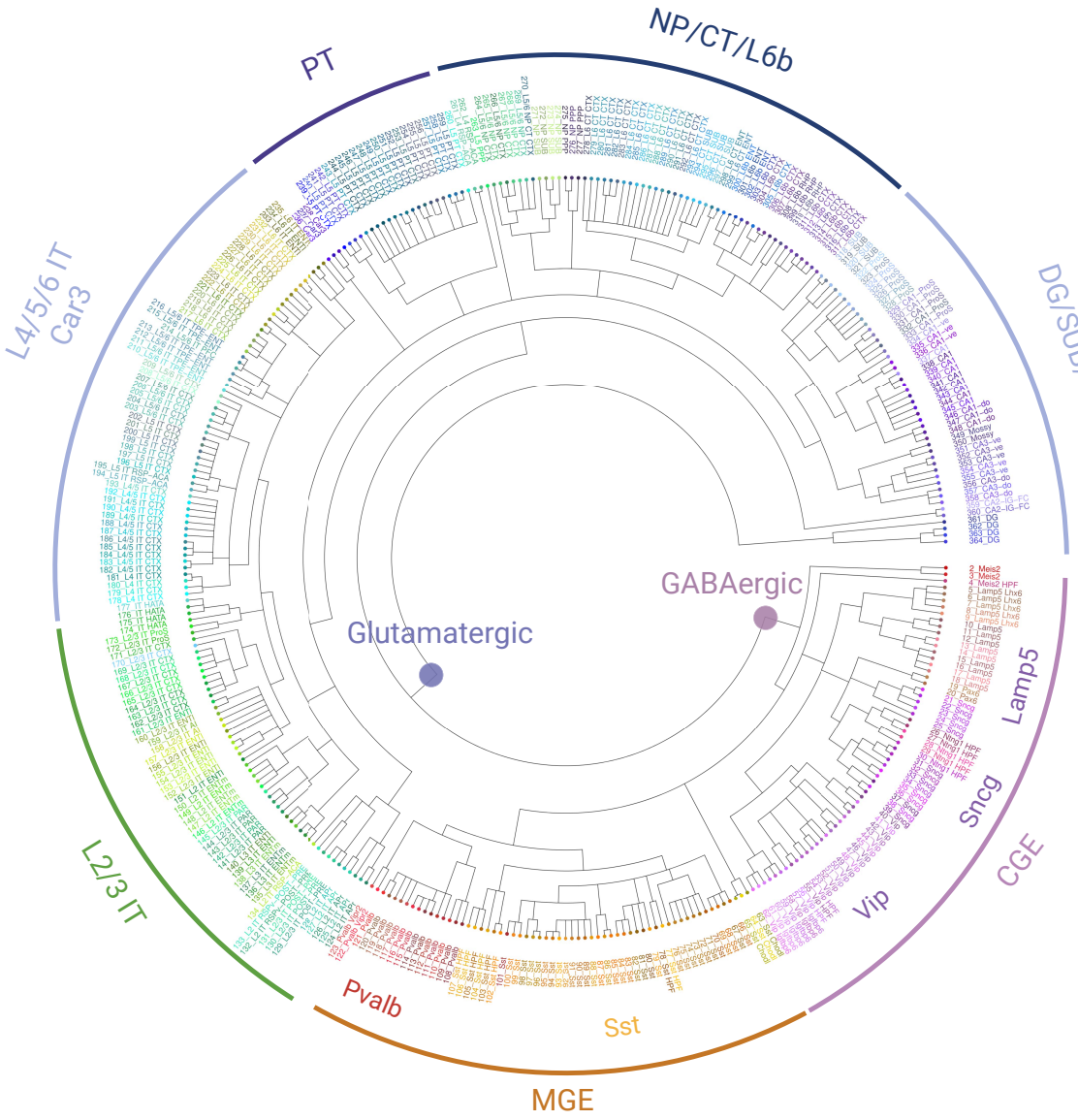

Figure 2.3: The hierarchical organisation of neuronal cell types. Hierarchical clustering of neurons based on their gene expression. Clusters are named according to cluster number, layer (if applicable), subclass, and area (if applicable); 25 non-neuronal and one glutamatergic cluster(s) are not shown to reduce visual clutter. CGE, caudal ganglionic eminence; MGE, medial ganglionic eminence; IT, intratelencephalic; PT, pyramidal tract; NP, near-projecting; CT, corticothalamic, DG, dentate gyrus; SUB: subiculum; CA, cornu ammonis. Taxonomy by Yao et al. (2021), created with R and Biorender.

This design increases throughput but may reduce the number of detected genes (Wang et al., 2021).

The efficiency of capturing RNA transcripts therefore depends on the experimental method, but no existing method can capture all of the RNA in a cell. Furthermore, transcription occurs in sporadic bursts rather than as a steady flow (Raj et al., 2006), such that some expressed genes will not be actively transcribed at the exact moment of sequencing. For these reasons, the resulting count matrix contains a large fraction of zeros corresponding to genes not detected in a particular cell. Some of these genes are indeed not expressed. Many others, however, failed to be captured or amplified, a phenomenon known as dropout (Stegle et al., 2015).

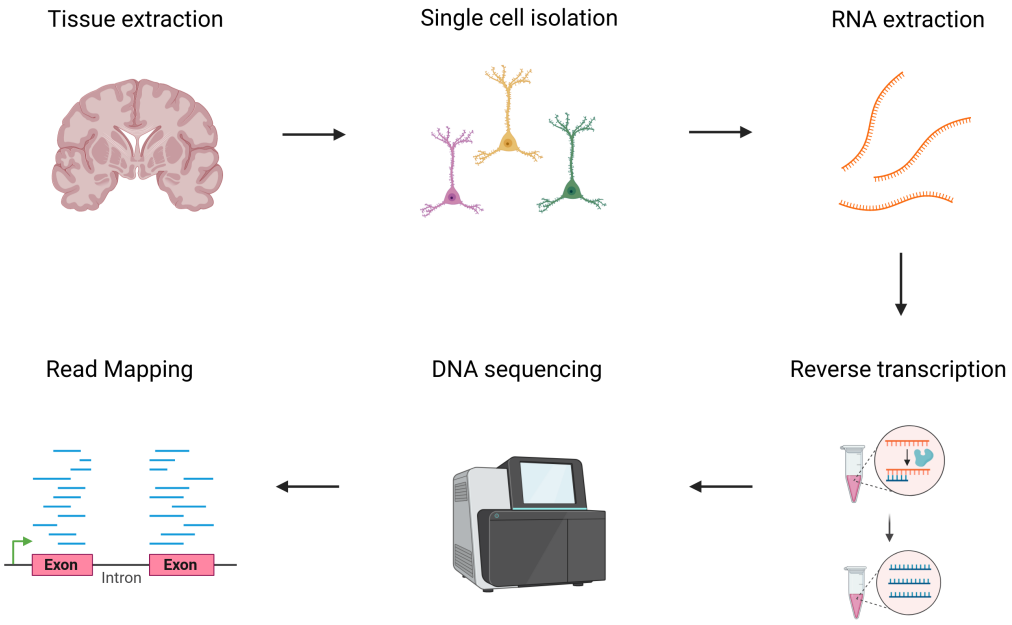

Figure 2.4: Workflow for generating scRNA-seq data. First, the brain area(s) of interest is extracted, from which individual cells or nuclei are isolated. Each cell's RNA is extracted, reverse-transcribed in the complementary DNA and sequenced in reads of a few hundred base pairs. Reads are then mapped to the reference genome, resulting in a cell-by-gene matrix of counts. Figure created using Biorender.

2.2.2 Analysis of scRNA-seq data involves multiple challenges

The experimentally collected RNA count matrix has several properties that prevent straightforward interpretation. The data is high dimensional because it contains at least as many variables (genes) as observations (cells or nuclei). It is noisy because even identical cells would be represented with slightly different gene counts due to dropout. Finally, the data is non-Gaussian because it consists of discrete counts with highly skewed distributions. Thus, widely used analysis tools such as linear regression and principal component analysis may be inappropriate—at least before preprocessing.

Single-cell data analysis therefore requires several steps (Luecken and Theis (2019) and Andrews et al. (2021); Fig. 2.5). The first step is quality control based on the joint distribution of several statistics, such as the number of reads per cell and the fraction of mitochondrial genes per cell.

After quality control, the data is preprocessed to correct for technical sources of variability (Vallejos et al., 2017). A simple and widely used approach is to normalise the number of counts per cell to a fixed number, such as counts per million. This assumes that the cells originally contained the same number of molecules, and the experiment introduced artificial variability, which has to be corrected. The data is also transformed to reduce its skewness, for example, by using a log-transform.

The next step is to filter out genes that are unlikely to be informative for downstream analysis. For instance, genes from which only a few transcripts were detected across all cells are usually removed. Genes with consistently high expression across cells are also excluded

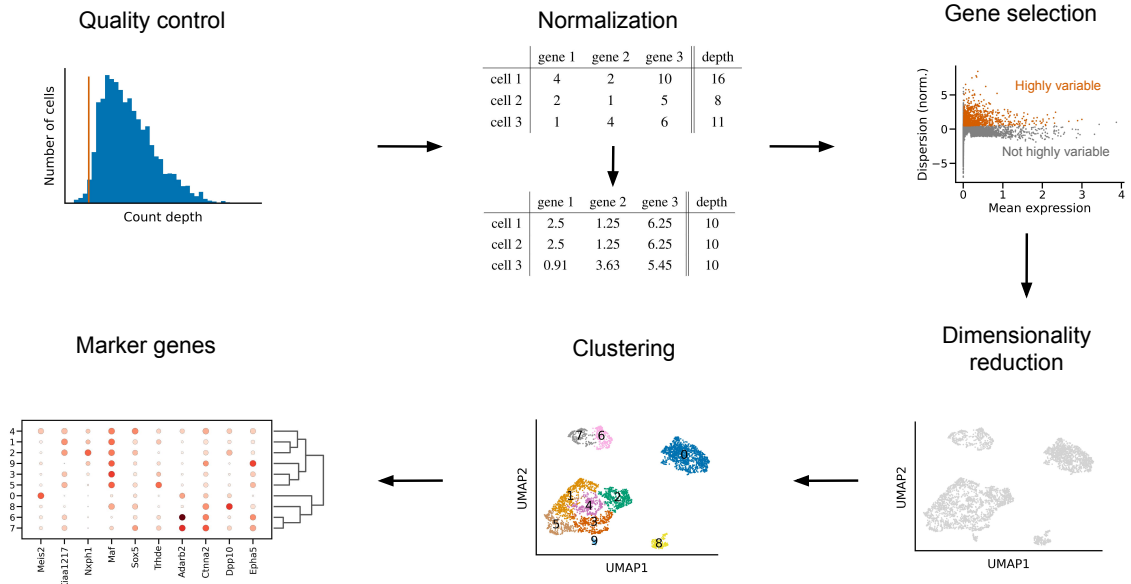

Figure 2.5: Workflow for analysing scRNA-seq data. First, cells are subjected to quality control to ensure that the counts from each cell do, in fact, derive from one and only one healthy cell. Next, the RNA counts are normalised for technical variation. This step also includes a variance-stabilising transformation. Next, gene selection and dimensionality reduction identify a parsimonious description of the biological variation. This enables the clustering of cells, which can identify cell types and states. Marker genes distinguish different clusters. Data from Colquitt et al. (Colquitt et al., 2021).

since they do not help to distinguish different cell types. Typically, several thousand genes with a high variability compared to mean expression across cells are selected (Macosko et al., 2015; Zheng et al., 2017).

The resulting dataset consists of high-quality cells and informative genes. Its dimensionality can be reduced further since the expression of different genes is often correlated, such that cells occupy a low-dimensional manifold within the space of all gene expression patterns. This manifold can be uncovered using dimensionality reduction. First, principal component analysis (PCA) determines 20 to 50 linear combinations of genes that capture most of the variability. Next, a nonlinear method like t-SNE (Van der Maaten and Hinton, 2008) or UMAP (McInnes et al., 2018) is used to find an even lower dimensional representation for visualising the data.

The PCA-transformed data is now amenable to clustering algorithms that would otherwise suffer from the curse of dimensionality. The goal of clustering is to organise cells into groups that are similar to each other but different from other cells (Kiselev et al., 2019). At a biological level, these groups should correspond to cell types or cell states. The nested nature of cell types (see below) makes hierarchical clustering methods attractive for producing small clusters nested within larger clusters.

Clustering is followed by identifying genes that statistically distinguish the clusters (Soneson and Robinson, 2018). If a differentially expressed gene happens to be a known marker gene, this identifies a previously unknown cluster as a particular cell type.

2.2.3 Characterising cell types requires many modalities

We introduced transcriptomics with the observation that cell types are multidimensional entities. Although transcriptome-wide variation correlates well with other dimensions of neuronal diversity, such as morphology, the agreement is not perfect (e.g., Kim et al. (2020) and Peng et al. (2021)). And even if different modalities aligned perfectly, we would still need to translate between them.

A powerful technique for the multimodal characterisation of cell types is Patch-seq, which performs whole-cell patch clamping followed by scRNA-seq to profile the transcriptomic and electrophysiological (and morphological) features of individual cells (Cadwell et al., 2016; Fuzik et al., 2016). PatchSeq has been used to define 28 "MET-types" of GABAergic interneurons with consistent morphological (M), electrophysiological (E) and transcriptomic (T) variation (Gouwens et al., 2020). These MET-types can function as a "Rosetta stone" for integrating multiple aspects of neuronal diversity (Lipovsek et al., 2021; Zeng, 2022).

Spatial information is another important addition to transcriptomic data (Lein et al., 2017; Burgess, 2019). Since scRNA-seq requires the dissociation of individual cells, it does not preserve their precise location within the dissected tissue. Novel methods for "spatial transcriptomcs" instead sequence (Rodriques et al., 2019; Stickels et al., 2021) or image (Chen et al., 2015a; Eng et al., 2019) gene expression *in situ*, i.e. in place. This opens up the exciting possibility of combining transcriptomics and connectomics (Chen et al., 2019; Chen et al., 2022) or in vivo activity recordings (Bugeon et al., 2022; Condylis et al., 2022). Spatial methods currently compromise between the number of cells, transcriptome coverage, sequencing depth and spatial resolution. This compromise can be mitigated by integrating the spatial data with large-scale, high-depth scRNA-seq datasets. Chapter 6 will use this approach to extrapolate insights from a spatial dataset.

3

Outline of the thesis

Despite extensive experimental and theoretical work, it is not fully resolved why the brain contains such diverse inhibitory cell types. We study this question in three separate but connected projects. We start out from a top-down, functional perspective, using machine learning to understand interneuron functions. Next, we compare this functional approach with one based on interneuron evolution and development. Finally, we study the behavioural relevance of interneuron diversity by comparing patterns of gene expression across species.

Interneuron diversity enables specific inhibition

A central function of interneurons is stabilizing cortical circuits by feedback inhibition. This function is mainly attributed to PV-expressing interneurons that synapse onto the perisomatic region of pyramidal cells. However, pyramidal cells also have apical dendritic compartments, which are innervated by SST-expressing interneurons. We propose that the coexistence of PV+ and SST+ types allows for distinct feedback loops tailored to different neuronal compartments. We test this idea using recent advances in training spiking network models. Specifically, we optimize interneuron connectivity and short-term plasticity to achieve compartment-specific inhibition of pyramidal neurons. Consistent with our hypothesis, the optimization transformed a homogeneous interneuron population into two distinct classes resembling PV+ and SST+ interneurons. Our gradient-based optimization of biologically realistic models provides a tool for exploring the relationship between the structure and function in cortical circuits.

This work is published in the following journal article and reproduced in Chapter 4 with minor modifications.

> **Keijser, J.** and Sprekeler, H., 2022. Optimizing interneuron circuits for compartment-specific feedback inhibition. *PLOS Computational Biology*, 18(4), p.e1009933.

Interneuron diversity predates some of its current functions

In our first project, we showed that a specific combination of short-term plasticity and synapse specificity enables compartment-specific feedback. But was compartment-specific inhibition indeed the function for which PV+ and SST+ cells originally evolved? Does the compartmental structure of pyramidal cells shape the diversification of PV+ and SST+ interneurons over development? Our second project investigated these questions using data on the development and evolution of PV+ and SST+ interneurons on one hand and pyramidal cells on the other. These data speak against the idea that the compartmental structure of pyramidal cells drove the diversification of PV+ and SST+ interneurons. For example, comparative single-cell RNA-seq and electrophysiology indicate that PV+ and SST+ cells, but not the dendritic physiology of pyramidal cells, existed in the last common ancestor of mammals and reptiles. Thus, PV+ and SST+ cells likely evolved (and developed) the properties enabling compartment-specific inhibition before there was selective pressure for this function. The properties of a cell type may, therefore, have been assembled during its meandering evolutionary history rather than being optimized for its current purpose.

This work is published in the following journal article and reproduced in Chapter 5 with minor modifications.

> **Keijser, J.** and Sprekeler, H., 2023. Cortical interneurons: fit for function and fit to function? Evidence from development and evolution. *Frontiers in Neural Circuits*, 17, p.1172464.

Homologous interneurons show functionally-relevant differences

Like many interneuron studies, the first parts of the thesis focused on a few broadly defined cell-type families. However, interneuron diversity encompasses additional cell types, each containing many subtypes with a unique pattern of gene expression. Recent work indicates that these gene expression patterns can predict how an interneuron's activity is modulated by the animal's internal state—at least in mice (Bugeon et al., 2022). Here, we leverage this link between transcriptomics and activity to study behaviourally relevant expression patterns across species. We expected that the homology of interneuron subclasses would be reflected in the conservation of transcriptomic patterns known to predict state modulation in mice. Contrary to this expectation, we only identified similar axes of gene expression between closely related species (mice and humans) but not between more distantly related species (turtles and zebra finches). Additionally, we compared the expression of cholinergic receptors, which are causally related to state modulation. We used these patterns to construct circuit models and understand how species-specific gene expression could lead to species-specific state modulation. These results suggest that, although interneurons are evolutionarily old, their state modulation may have evolved in a lineage-specific manner.

This work is available as a preprint and reproduced in Chapter 6 with minor modifications.

Keijser, J., Hertäg, L. & Sprekeler, H., 2023. Transcriptomic correlates of state modulation in GABAergic interneurons: A cross-species analysis. *bioRxiv*, pp.2023-12.

Additional publications

In addition to the aforementioned work, I contributed to two projects that were published in the following journal articles.

Naumann, L.B., **Keijser, J.** and Sprekeler, H., 2022. Invariant neural subspaces maintained by feedback modulation. *Elife*, 11, p.e76096.

Schroeder, A., Pardi, M.B., **Keijser, J.**, Dalmay, T., Groisman, A.I., Schuman, E.M., Sprekeler, H. and Letzkus, J.J., 2023. Inhibitory top-down projections from zona incerta mediate neocortical memory. *Neuron*, 111(5), pp.727-738.

4

Optimizing interneuron circuits for compartment-specific feedback inhibition

4.1 Context within thesis

The widely observed balance of excitation and inhibition is classically attributed to parvalbumin (PV)-positive interneurons that synapse onto the perisomatic region of excitatory cells. However, excitatory pyramidal cells have other compartments, such as apical dendrites, which are regulated by somatostatin (SST)-positive interneurons. This chapter investigates how PV+ and SST+ interneurons complementarily balance excitation and inhibition in both pyramidal somata and apical dendrites. We show that optimizing a network's connectivity and short-term plasticity for compartment-specific inhibition causes interneurons to diversify into discrete populations, resembling PV+ and SST+ cells. However, the optimization can also create a continuum of diversity. Although compartment-specific inhibition is therefore enabled by PV/SST-diversity, the existence of alternative solutions points to a key challenge for optimization-based models, as discussed in Chapter 7.

A possible interpretation of our model is that the need for compartment-specific feedback explains the existence of PV+ and SST+ interneurons or at least some of their features. We will investigate this interpretation in the next chapter (Ch. 5).

4.2 Abstract

Cortical circuits process information by rich recurrent interactions between excitatory neurons and inhibitory interneurons. One of the prime functions of interneurons is to stabilize the circuit by feedback inhibition, but the level of specificity on which inhibitory feedback operates is not fully resolved. We hypothesized that inhibitory circuits could enable separate feedback control loops for different synaptic input streams by means of specific feedback inhibition to

different neuronal compartments. To investigate this hypothesis, we adopted an optimization approach. Leveraging recent advances in training spiking network models, we optimized the connectivity and short-term plasticity of interneuron circuits for compartment-specific feedback inhibition onto pyramidal neurons. Over the course of the optimization, the interneurons diversified into two classes that resembled parvalbumin (PV) and somatostatin (SST) expressing interneurons. Using simulations and mathematical analyses, we show that the resulting circuit can be understood as a neural decoder that inverts the nonlinear biophysical computations performed within the pyramidal cells. Our model provides a proof of concept for studying structure-function relations in cortical circuits by combining gradient-based optimization and biologically plausible phenomenological models.

4.3 Introduction

Cortical inhibitory interneurons vary dramatically in shape, gene expression pattern, electrophysiological and synaptic properties and in their downstream targets (Tremblay et al., 2016). Some cell types, e.g., somatostatin (SST)-positive interneurons (Urban-Ciecko and Barth, 2016) and some neurogliaform cells in layer 1 (Abs et al., 2018), predominantly project to pyramidal cell (PC) dendrites. Others—e.g., parvalbumin-positive (PV) basket and chandelier cells—primarily inhibit the peri-somatic domain of PCs (Hu et al., 2014). Some interneurons receive depressing synapses from PCs, others facilitating synapses (Markram et al., 1998; Reyes et al., 1998). But what is the function of these differences?

One of inhibition's core functions is to prevent run-away excitation (Isaacson and Scanziani, 2011) by means of feedback inhibition that tracks excitatory inputs. This has led to the concept of excitation-inhibition (E/I) balance (Van Vreeswijk and Sompolinsky, 1996), i.e., the idea that strong excitatory currents are compensated by inhibitory currents of comparable size. E/I balance is thought to shape cortical dynamics (Van Vreeswijk and Sompolinsky, 1996; Brunel, 2000) and computations (Hennequin et al., 2014; Rubin et al., 2015) and can be established by means of inhibitory forms of plasticity (Vogels et al., 2011; Mackwood et al., 2021; Hennequin et al., 2017). Selective disruptions of E/I balance are thought to play a key role during learning (Letzkus et al., 2015), while chronic disturbances have been implicated with psychiatric diseases, including autism (Yizhar et al., 2011; Sohal and Rubenstein, 2019) and schizophrenia (O'Donnell, 2011; Grent et al., 2018).

Originally conceived as a balance on average (Van Vreeswijk and Sompolinsky, 1996), E/I balance turned out to be specific to sensory stimuli (Wehr and Zador, 2003; Rupprecht and Friedrich, 2018), in time (Okun and Lampl, 2008; Renart et al., 2010), across neurons (Xue et al., 2014) and to neural activation patterns (Bhatia et al., 2019). The number of excitatory and inhibitory synapses could even be balanced at the subcellular level (Iascone et al., 2020), in a cell-type specific way (Karimi et al., 2020). Given this high specificity, we hypothesized that excitation and inhibition also balance individually in different neuronal compartments and that this could be mediated at least in part by compartment-specific feedback inhibition.

Different neuronal compartments often receive input from different sources (Petreanu et al., 2009) and integrate these inputs nonlinearly by means of complex cellular dynamics (London and Häusser, 2005; Poirazi and Papoutsi, 2020). For example, the apical dendrites of L5

pyramidal cells (PCs) can generate nonlinear calcium events in response to coincident somatic and dendritic inputs (Larkum et al., 1999). Hence, neuronal output spike trains can have a complex nonlinear dependence on the inputs arriving in different compartments. This poses a challenge for compartment-specific feedback inhibition, which would require interneurons to invert the nonlinear dependence by recovering local dendritic input from the pyramidal output. It is, therefore, far from clear that a compartment-specific feedback inhibition can be achieved at all by means of biologically plausible circuits. If it can, however, it would have to rely on an interneuron circuit that is closely matched to the electrophysiological properties of the cells it inhibits. Parts of the complexity of cortical interneuron circuits could then be interpreted in light of the intrinsic properties of PCs.

Unfortunately, the nature of such a correspondence between the electrophysiology of inhibited cells and suitable interneuron circuits is far from obvious. We reasoned that we could gain insights by means of a model-based optimization approach, in which interneuron circuits are optimized for feedback inhibition onto pyramidal cells with given biophysical properties. Here, we illustrate this ansatz by optimizing interneuron circuits for a nonlinear two-compartment model of L5 pyramidal cells (Naud et al., 2014). We show that over the course of the optimization, an initially homogeneous interneuron population diversifies into two classes, which share many features of cortical PV and SST interneurons. One class primarily inhibits the somatic compartment of the PCs and receives depressing synaptic inputs. The other class primarily inhibits PC dendrites and receives facilitating inputs. We use further computer simulations and mathematical analyses to investigate the mechanism underlying this interneuron diversification. We show how the diversification can be understood from an encoding-decoding perspective, in which the biophysics of the PCs encode two different input streams in a multiplexed code (Naud and Sprekeler, 2018), which is in turn decoded by the interneuron circuit. Finally, we identify several factors that determine if interneurons fall into discrete cell types or exist along a continuum. Together, these findings support the idea that parts of the complexity of cortical interneuron circuits could be interpreted in light of the intrinsic properties of PCs and illustrate how modelling could provide a means of unravelling these interdependencies between the cellular and the circuit level.

4.4 Results

To investigate which aspects of cortical interneuron circuits can be understood from the perspective of compartment-specific inhibition, we studied a spiking network model comprising pyramidal cells (PCs) and interneurons (INs) (see Methods). PCs were described by a two-compartment model consisting of a soma and an apical dendrite. The parameters of this model were previously fitted to capture dendrite-dependent bursting (Naud et al., 2014). PCs received time-varying excitatory inputs in both the somatic and the dendritic compartment and inhibitory inputs from INs. The excitatory inputs to both compartments consisted of alternating currents of varying amplitude. The two currents were initially uncorrelated; we will investigate the effect of this assumption later. INs were described by an integrate-and-fire model. They received excitatory inputs from the PCs, and inhibitory inputs from other INs.

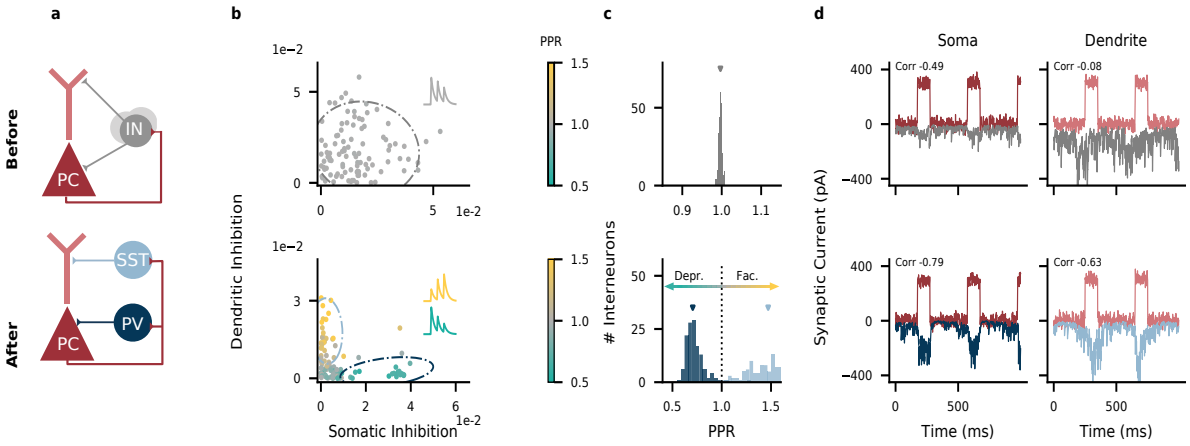

Figure 4.1: Interneuron diversity emerges in networks optimized for compartment-specific inhibition. (a) Network structure before (top) and after optimization (bottom). PC, pyramidal cell; IN, interneuron; PV, parvalbumin-positive IN; SST, somatostatin-positive IN. Recurrent inhibitory connections among INs were omitted for clarity. **(b)** Strength of somatic and dendritic inhibition from individual INs. Dashed lines: 95% density of a Gaussian distribution (top) and a mixture of two Gaussian distributions (bottom) fitted to the connectivity and Paired Pulse Ratio (PPR) data of 5 networks (marginalized over PPR). **(c)** PPR distribution (data from 5 networks). Mean PPR before optimization: 1.00; after optimization: 0.73 (PV cluster, $n = 133$) and 1.45 (SST cluster, $n = 113$). **(d)** Excitatory (red) and inhibitory (top: grey, bottom: blue) currents onto PC compartments (average across $N_E = 400$ PCs). The excitatory inputs to the two compartments are uncorrelated. Inset: correlation between compartment-specific excitation and inhibition.

We optimized the interneuron circuit for compartment-specific feedback inhibition. In the presence of time-varying external input, feedback inhibition tracks excitatory inputs in time (Van Vreeswijk and Sompolinsky, 1996; Renart et al., 2010). We therefore enforced compartment-specific feedback inhibition by minimizing the mean squared error between excitatory and inhibitory inputs in both compartments by means of gradient descent with surrogate gradients (Zenke and Ganguli, 2018). Importantly, we optimized not only the strength of all synaptic connections in the network but also the short-term plasticity of the PC → IN connections (see Methods).

Interneuron diversity emerges during optimization

Before the optimization, interneurons formed a single, homogeneous group (Fig. 4.1a, top). Most inhibited both somatic and dendritic compartments (Fig. 4.1b, top) and PC → IN connections showed non-specific synaptic dynamics (Fig. 4.1c, top). Synaptic dynamics were quantified using the paired-pulse ratio (PPR), the relative amplitude of consecutive postsynaptic potentials (see Methods). A PPR smaller than 1 indicates that later postsynaptic potentials are weaker and therefore corresponds to short-term depression. A PPR larger than 1 corresponds to short-term facilitation. Excitation and inhibition were poorly correlated, particularly in the dendrite (Pearson correlation coefficients 0.49 (soma) & 0.08 (dendrite)), suggesting that the network did not generate compartment-specific feedback inhibition (Fig. 4.1d, top). The relatively high correlation between somatic excitation and inhibition is explained by the fact that, in a recurrent network, inhibition is bound to track excitation to some extent (Van Vreeswijk and Sompolinsky, 1996; Renart et al., 2010).

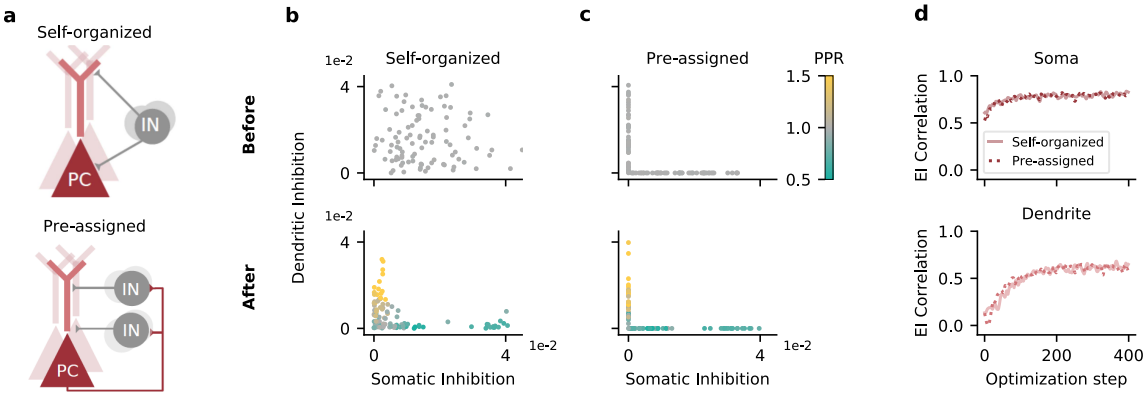

Figure 4.2: Compartment-assigned interneurons develop into PV- and SST-like populations. (a) Circuit before learning. Top, interneurons (INs) can inhibit both compartments of principal cells (PCs) and need to self-organize, as in Fig. 1. Bottom, INs are pre-assigned to inhibit a single PC compartment. **(b)** IN→PC weights before (top) and after (bottom) optimization. Interneurons self-organize into a population that preferentially inhibits the soma and a population that preferentially inhibits the dendrites. Data differs from Fig. 4.1 due to random parameter initialization and sampling of training data. **(c)** As (b), but with interneurons randomly assigned to inhibit a single compartment (soma or dendrite). Mean PPR: 0.72 (soma-inhibiting population), 1.17 (dendrite-inhibiting population). **(d)** Correlation between compartment-specific excitation and inhibition over the course of the optimization. Solid line: INs were not assigned to a single compartment (Self-organized). Dashed line: INs were assigned to a single compartment (Pre-assigned). Data is smoothed with a Gaussian kernel (width: 2).

During optimization, the interneurons split into two groups (Fig. 4.1a, bottom) with distinct connectivity (Fig. 4.1b, bottom; see also Connectivity among interneurons) and short-term plasticity (Fig. 4.1c, bottom). One group received short-term depressing inputs from PCs and preferentially targeted their somatic compartment, akin to PV interneurons. The other group received short-term facilitating inputs from PCs and targeted their dendritic compartment, akin to SST interneurons. For simplicity, we will henceforth denote the two interneuron groups as PV and SST interneurons. After the optimization, excitation and inhibition were positively correlated in both compartments (Pearson correlation coefficients 0.79 (soma) & 0.63 (dendrite); Fig. 1d, bottom). Note that the E/I balance is slightly less tight in time in the dendrites than in the somata (Fig. 4.1d), because synaptic short-term facilitation causes a delay in the signal transmission between PCs and SST interneurons (Pouille and Scanziani, 2004) (more details below).

To confirm the benefit of two non-overlapping interneuron classes, we performed control simulations in which each interneuron was pre-assigned to target either the soma or the dendrite while synaptic strengths and short-term plasticity were optimized. Consistent with the benefit of specialization, the correlation of excitation and inhibition in the two compartments was as high as in fully self-organized networks (Fig. 4.2). Optimized networks with pre-assigned interneuron classes also showed the same diversification in their short-term plasticity, resembling that of PV and SST neurons (Figs. 4.2, Fig. A.1).

Feedback inhibition decodes compartment-specific inputs

For compartment-specific feedback inhibition, the interneuron circuit has to retrieve the somatic and dendritic input to PCs from the spiking activity of the PCs. This amounts to inverting

the nonlinear integration performed in the PCs (Fig. 4.3a). How does the circuit achieve this? Recently, it was proposed that the electrophysiological properties of PCs support a multiplexed neural code that simultaneously represents somatic and dendritic inputs in temporal spike patterns (Naud and Sprekeler (2018), Fig. 4.3b). In this code, somatic input increases the number of events, where events can either be single spikes or bursts (see Methods). Dendritic input, in turn, increases the probability that a somatic spike is converted into a burst (burst probability). Providing soma- or dendrite-specific inhibition then amounts to decoding the event rate or burst probability, respectively. Such decoding can be achieved in circuits with short-term plasticity and feedforward inhibition (Naud and Sprekeler, 2018), and we expected that our network would arrive at a similar decoding scheme.

We tested this hypothesis by injecting current pulses to PC somata and dendrites (see Methods). Stronger dendritic input increased the burst probability, which increased the firing rate of SST interneurons via facilitating synapses. The increased SST rate increased dendritic inhibition (Fig. 4.3c-e, top). Analogously, stronger somatic input increased the event rate, which increased the firing rate of PV interneurons via depressing synapses. The increased PV rate increased somatic inhibition (Fig. 4.3c-e, bottom). Importantly, inhibition was specific to each compartment (shaded lines indicate input strength to the other compartment): Because PV interneurons were selectively activated by PC events, somatic inhibition was largely unaffected by dendritic excitation. Similarly, SST interneurons were selectively activated by PC bursts, such that dendritic inhibition was largely unaffected by somatic excitation. In the model, interneurons therefore provide compartment-specific inhibition by demultiplexing the neural code used by the PCs.

In networks trained without short-term plasticity, SST neurons were not selectively activated by bursts, and therefore dendritic inhibition did not balance dendritic excitation (Figs. 4.3f, A.2c). A soma-specific E/I balance also required short-term plasticity, but only for weak somatic excitation—consistent with the relatively high somatic E/I correlation at the start of training (Fig. 4.1). In the networks trained without short-term plasticity, the multiplexed neural code was unaltered because the biophysics of the PCs are the same, but the decoding by the interneuron circuit fails, most prominently for the dendritic input. In our model, short-term plasticity is therefore necessary for compartment-specific feedback inhibition.

Conditions for the emergence of discrete interneuron classes

PV and SST neurons largely form two non-overlapping cell types (Harris et al., 2018; Tasic et al., 2018; Scala et al., 2020), but in our model, they can also exist along a continuum. This depends on three model parameters. First, the correlations between compartment-specific inputs. So far, we assumed that PC somata and dendrites receive uncorrelated input. But recent work suggests that somatic and dendritic activity are correlated (Beaulieu-Laroche et al., 2019; Francioni et al., 2019), potentially reducing the need for compartment-specific inhibition. We therefore tested how correlated inputs affect interneuron specialization by optimizing separate networks for different input correlations. We found that increasing correlation between somatic and dendritic inputs gradually reduced the separation between the interneuron classes (Fig. 4.4a,b). For high input correlation, optimized networks contained a continuum in their connectivity and short-term plasticity (Fig. 4.4a,b). However, the presence of short-term plasticity was necessary

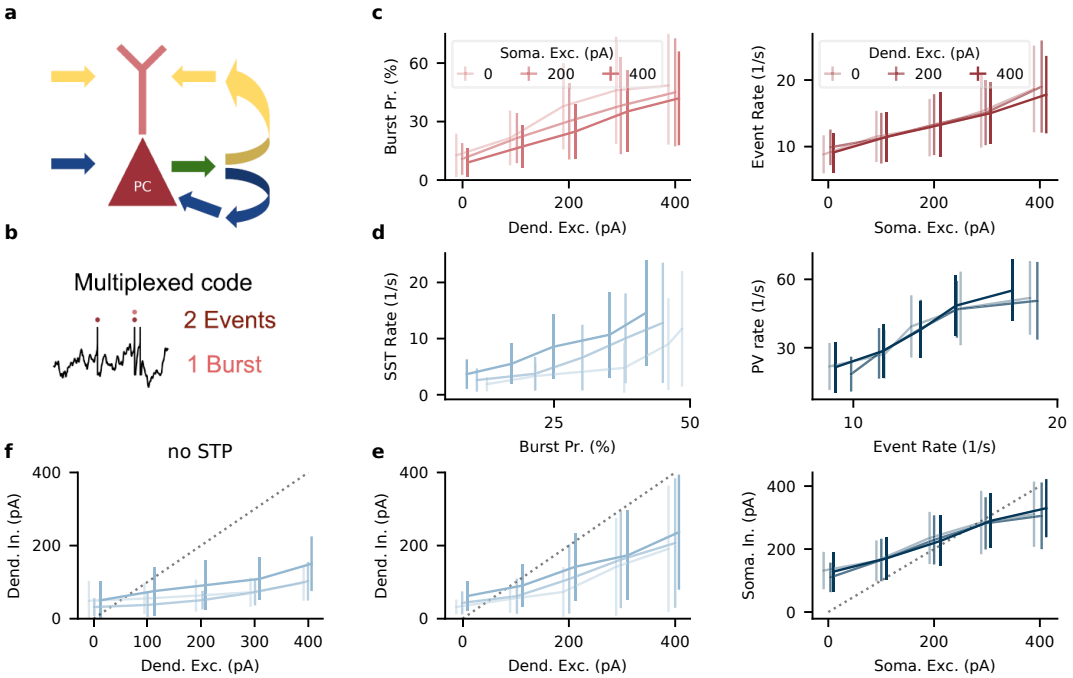

Figure 4.3: The interneuron circuit decodes somatic and dendritic inputs to PCs. (a) PC somata and dendrites receive uncorrelated input streams (yellow and blue) that, from PC output spikes (green), have to be separated into compartment-specific inhibition (yellow and blue). **(b)** PCs use a multiplexed neural code. Somatic input leads to events (singlets or bursts). Dendritic input converts singlets into bursts. **(c)** Left: Excitatory input to PC dendrites increases burst probability. In this and other top panels (d,e), the shading indicates the strength of background somatic input. Right: Excitatory input to PC somata increases event rate. In this and other bottom panels (d,e), the shading indicates the strength of background dendritic input. **(d)** Left: SST rate increases with bursts probability. Right: PV rate increases with PC events. **(e)** Left: dendritic inhibition increases with dendritic excitation, but is only weakly modulated by somatic excitation. Positions on x-axis are shifted by 10 pA for visual clarity; error bars indicate sd during 10 stimulus repetitions. Right: somatic inhibition increases with somatic excitation but is invariant to dendritic excitation. Dashed lines correspond to excitation = inhibition. **(f)** In networks trained without short-term plasticity, dendritic inhibition shows a weaker dependence on dendritic excitation and a stronger dependence on somatic excitation.

for a dendritic E/I balance for a range of input correlations (Fig. 4.4c). At high correlations, somatic and dendritic inputs are sufficiently similar to make the effect of short-term facilitation negligible. Note that although, in this case, distinct interneuron populations were not necessary, the presence of IN classes was also not harmful to E/I balance. A pre-assignment of the interneurons into classes maintained the E/I correlation in both compartments and for any correlation level (Fig. A.1).

Interneuron specialization also degraded with increasing baseline activity of the INs (Fig. A.2), because high firing rates allow non-specialized inhibition to be cancelled by equal and opposite disinhibition (see mathematical analysis in appendix, Computational Methods). The dependence on baseline activity results from the fact that disinhibition is limited by how much the firing of the interneurons can be reduced. In that regard, it is not the baseline firing rate itself that determines the specialization – which is often higher for interneurons than for PCs (see, e.g., Yu et al. (2019)) – but the relation between the baseline and the dynamic range of the firing rates that is required for the appropriate disinhibition. Note also that a pre-assignment of interneurons into classes again maintained the E/I correlation for different

baseline activity levels (Fig. A.1).

Finally, interneuron specialization was reduced in networks with heterogeneous inhibitory connectivity. So far, we used homogeneous IN→PC connectivity, i.e., each IN inhibited all PC somata with the same strength and all PC dendrites with the same strength. In simulations in which interneurons were free to inhibit each soma (and each dendrite) with a unique strength, PV and SST clusters also emerged, but we additionally observed non-specialized interneurons (Fig. A.4). But do these non-specialized interneurons play an active role in the computation, or are they not necessary and therefore left behind by gradient descent once the problem is solved? The fact that the E/I correlation is not higher than for the homogeneous setting suggests the latter (heterogeneous networks: 0.844 +/- 0.011 (soma) and 0.702 +/- 0.022 (dendrite); homogeneous networks: 0.842 +/- 0.006 (soma) and 0.717 +/- 0.014 (dendrite)). In sum, a compartment-specific E/I balance seems to require a diversity of interneurons, but the degree to which the interneurons fall into discrete classes depends on a variety of factors.

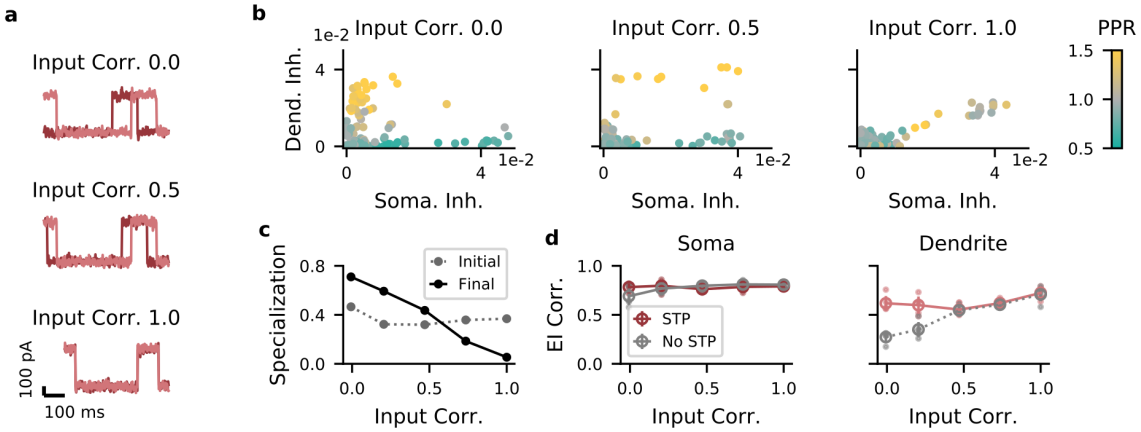

Figure 4.4: Correlations between dendritic and somatic input reduce interneuron specialization. (a) Examples for synaptic traces corresponding to different correlation levels. Dark red, somatic current; light red, dendritic input. **(b)** Strength of somatic vs. dendritic inhibition from all INs. Left, middle, right: input correlation coefficient 0 (low), 0.5 (medium), and 1 (high), respectively. **(c)** Specialization of IN → E weights. If each IN targets either soma or dendrites, the specialization is 1 (see Methods). Gray: specialization of the initial random network; black: specialization after optimization. **(d)** Left: In the soma, excitation and inhibition are balanced across a broad range of input correlations, with or without short-term plasticity (STP). Right: In the dendrites, excitation and inhibition are balanced only with STP when input correlations are small.

Connectivity among interneurons

Because interneurons subtypes also differ in their connectivity to other interneurons (Pfeffer et al., 2013; Jiang et al., 2015), we included IN → IN synapses in our optimization. After classifying INs as putative PV and SST neurons using a binary Gaussian mixture model, we found that the connections between the interneuron classes varied systematically in strength. While PV ↔ PV connections, PV → SST connections and SST ↔ SST connections were similar in strength on average, SST → PV were consistently stronger (Fig. 4.5a), presumably to compensate for the relatively low SST rates (Fig. 4.3d).

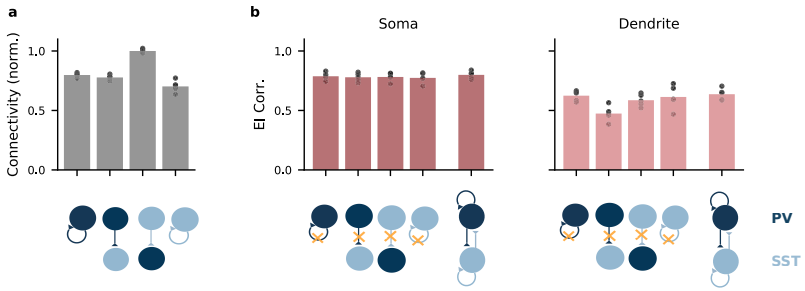

Figure 4.5: Recurrent inhibitory connectivity after learning. (a) Connectivity between IN populations. From left to right: PV↔PV, PV→SST, SST→PV, SST↔SST. Bars indicate mean over all networks; dots indicate individual networks. **(b)** Performance as measured by the correlation between excitation and inhibition to PC soma (left) and dendrites (right) of networks optimized while lacking specific connections. Data at the very right: E/I correlation in network with unconstrained connectivity. Only loss of PV → SST connectivity during optimization has a clear effect on dendritic E/I correlations after optimization. Open circles, mean over 5 batches of 8 stimuli with random amplitudes. Small filled circles, individual batches.

To investigate which connections were necessary, we simulated knockout experiments in networks with pre-assigned interneuron classes, in which we removed individual connection types prior to optimization. We found that only PV → SST connections were necessary for a dendritic E/I balance (Fig. 4.5b). Note that although earlier work did not find PV → SST connectivity in the primary visual cortex of young mice (Pfeffer et al., 2013), these connections seem to be present in the primary visual and somatosensory cortex of older animals (Jiang et al., 2015; Karnani et al., 2016b).

To understand the role of the different IN→IN connections, we performed a mathematical analysis of a simplified network model. The model also contains a population of principal cells (PC) and two populations of interneurons corresponding to PV and SST interneurons, but in contrast to the spiking model, neural activities are represented by continuously varying rates. The population rates of PV and SST interneurons are denoted by p and s, respectively. The activity of PCs is described by two rates: an event rate e that is driven by somatic input and a burst rate b that is driven by dendritic input. The burst rate is assumed to be independent of the somatic input, which is different from BAC firing (Larkum et al., 1999), but generates a linear model that is analytically tractable. The short-term plasticity of a given synapse type is characterized by a single, static parameter, which characterizes the relative efficiency at which events and bursts are transmitted. Synapses for which this parameter is 1 transmit events but not bursts, i.e., they are "perfectly depressing
. Synapses for which this parameter is 0 transmit only bursts, i.e., they are "perfectly facilitating". These assumptions allowed us to mathematically analyze the interneuron connectivity required for compartment-specific feedback inhibition. We will only summarize the results; the full analysis is described in Supporting information to Chapter 4.

Let us first consider the case of dendritic feedback inhibition. The model states that the activity s of the SST neurons is given by a linear combination of the event and burst rate: $s = A\,e + B\,b$, with factors A, B that depend on the connectivity and short-term plasticity in the circuit in a complicated way. If we assume that SST interneurons target exclusively the PC dendrites, compartment-specific feedback inhibition requires that the activity of SST interneurons depends on dendritic but not somatic input to PCs. Because those two inputs

drive the burst rate and event rate, respectively, this condition reduces to the mathematical condition that $A = 0$. Using the dependence of A on the circuit parameters (see Supporting information to Chapter 4), we get the condition

$$\beta W^{SST \leftarrow PC} - \alpha W^{SST \leftarrow PV} W^{PV \leftarrow PC} = 0\,, \tag{4.1}$$

where $W^{Y \leftarrow X}$ denotes the strength of the synaptic connection between population X and Y. The two parameters α, β are the short-term plasticity parameters and quantify how well events are transmitted via the PC→PV and PC→SST connections, respectively.

Condition [4.1] has an intuitive interpretation. The first term describes how much somatic PC input influences SST activity via the monosynaptic pathway PC → SST. The second term corresponds to the disynaptic pathway PC → PV → SST. The condition therefore states that unless PC→ SST connections are "perfectly facilitating" ($\beta = 0$), the disynaptic PC → PV → SST pathway is necessary (Fig. 4.5) to avoid that somatic input generates dendritic inhibition. The observation that a knock-out of these connections reduces the dendritic E/I correlation in the spiking network (Fig. 4.5b) can therefore be understood as a result of an imperfect facilitation in the PC→SST connection. Indeed, we observed that the synapses in the optimized spiking network are not perfectly facilitating. In fact, the Tsodyks-Markram model (Tsodyks et al., 1998) we used to describe the short-term plasticity in the spiking network cannot achieve perfect facilitation. In the presence of ongoing activity, even for an initial release probability $U = 0$, preceding spikes always leave behind a residual level of synaptic facilitation.

An analogous analysis suggests that disynaptic PC → SST→ PV inhibition is necessary to prevent dendritic inputs from generating somatic inhibition (Supporting information to Chapter 4), providing a possible function of experimentally observed SST → PV connectivity. At first sight, this appears in conflict with the observation that a knock-out of this connection did not reduce the E/I balance in the soma. However, because bursts are comparatively rare (Naud and Sprekeler, 2018), event rate and overall firing (including additional spikes in bursts) are highly correlated. Therefore, the overall firing rate is a good proxy for somatic input and imperfections in synaptic depression in the PC→PV connection do not introduce a sufficiently large problem to necessitate feedforward inhibition via the PC→SST→PV pathway.

Discussion

Feedback inhibition ensures the stability of cortical circuits (Tsodyks et al., 1997; Ozeki et al., 2009; Rubin et al., 2015; Sanzeni et al., 2020). Our model indicates that this feedback could operate on a level as fine-grained as different cellular compartments receive different input streams and that the required circuitry bears similarity to the one observed in cortex. In particular, we found that an optimization for feedback inhibition led to the emergence of two inhibitory cell classes that resemble PV and SST interneurons in their connectivity and short-term plasticity. This diversification was robust to correlations between somatic and dendritic input, although increasing correlations prompted the SST-like model neurons to contact not only the dendritic but also the somatic compartment. This is consistent with the extensive branching of cortical SST neurons within the layer that contains their cell body

(Urban-Ciecko and Barth, 2016). Even when the gradient-based optimization did not drive a clear division into cell classes, an artificial pre-assignment of the interneurons did not impair the feedback inhibition.

Specificity of feedback inhibition

We would like to emphasize that while we optimized for feedback inhibition in different neuronal compartments, the model operates on an ensemble level in the sense that all neurons in the network received the two same time-varying signals in their soma and dendrite. This allows the interneurons to use event or burst rates of the whole ensemble to infer somatic and dendritic inputs with high temporal fidelity (Naud and Sprekeler, 2018). The question of the specificity of feedback inhibition on the population level is an orthogonal one and not fully resolved. The dense and seemingly unspecific connectivity of many interneurons (Packer and Yuste, 2011; Fino et al., 2013) suggests that feedback inhibition operates on the level of the local population, blissfully ignoring the functional identity of the neurons it targets (Harris and Mrsic-Flogel, 2013). More recent results have indicated a correlation between the sensory tuning and the synaptic efficacy of interneuron-pyramidal cell connections, however, suggesting that feedback inhibition could operate on the level of functionally identified ensembles (Znamenskiy et al., 2018; Mackwood et al., 2021). A natural extension of this work would be to endow the pyramidal cells with tuning to different somatic and dendritic input streams and thereby define functional ensembles. Notably, the ensemble affiliation of a given neuron may differ for soma and dendrite, e.g., two populations of neurons could receive distinct somatic but identical dendritic inputs. How this would be reflected in the associated feedback-optimized interneuron circuit is an interesting question but beyond the scope of the present work.

Origins of interneuron diversity

A natural question for optimization-based approaches is how the optimization can be performed by biologically plausible mechanisms. The gradient-based optimization we performed relies on surrogate gradients (Zenke and Ganguli, 2018; Neftci et al., 2019) and a highly non-local backpropagation of errors both through the network and through time (Rumelhart et al., 1986; Werbos, 1990), mechanisms that are unlikely implemented verbatim in the circuit (Lillicrap and Santoro, 2019). We think of the suggested optimization approach rather as a means to understanding functional relations between different features of neural circuits, i.e., the relation between the biophysics of pyramidal cells and the surrounding interneuron circuits. At this point, we prefer to remain agnostic about the mechanisms that establish these relations. While an activity-dependent refinement of the circuit is likely, the diversification of the interneurons into PV and SST neurons is clearly not driven by activity-dependent mechanisms alone (Butt et al., 2005; Lim et al., 2018b; Mi et al., 2018). For example, SST Martinotti cells migrate to the embryonic cortex via the marginal zone, while PV basket cells migrate via the subventricular zone (Lim et al., 2018b). Their identity is hence determined long before they are integrated into functional circuits. These developmental programs are likely old on evolutionary time scales given that interneuron classes seem more conserved than pyramidal cell classes (Tosches et al., 2018; Colquitt et al., 2021). We therefore do not expect our mathematical optimization to mimic the evolutionary or developmental processes that generated interneuron diversity.

Experimental predictions

Given these considerations, we refrain from predictions regarding the optimization process. Still, the model can make predictions regarding the nature of the optimized state. First, it predicts that excitation and inhibition are balanced on short-time scales in both somatic and dendritic compartments. Second, it predicts that PV and SST rates correlate primarily with somatic and dendritic activity, respectively. In our model, this correlation is a consequence of the decoding that underlies the balance (Fig. 4.4), and it is experimentally more accessible than the underlying excitatory and inhibitory currents. Third, inhibiting SST neurons should increase PC bursting, as observed in hippocampus (Royer et al., 2012) and cortex (Gentet et al., 2012). The role of short-term facilitation could be tested by silencing the necessary gene Elfn1 (Sylwestrak and Ghosh, 2012; Stachniak et al., 2019). On a higher level, the model suggests a relation between the biophysical properties of excitatory neurons and the surrounding interneuron circuit. This is consistent, e.g., with the finding that the prevalence of pyramidal cells and dendrite-targeting Martinotti cells seems to be correlated across brain regions (Scala et al., 2019). This and other correlations between interneuron and pyramidal properties could be investigated systematically using recent electrophysiological and genomic data from different areas and organisms (Gouwens et al., 2020; Berg et al., 2021; Beaulieu-Laroche et al., 2021).

Model limitations and extensions

While the synaptic targets and the incoming short-term plasticity of the two emerging interneuron classes are similar to those of PV and SST interneurons, the optimized inhibitory circuitry is not a perfect image of cortex. Aside from the obvious incompleteness in terms of other interneuron types, other features, such as the often observed weak connectivity from PV to SST neurons (Pfeffer et al., 2013) did not result from the optimization (Fig. 4.5). Our approach also did not explain further cell type-specific electrophysiological properties: Exploratory simulations indicated that optimizing membrane time constants and spike-rate adaption parameters does not further improve the compartment-specific E/I balance. However, even if our assumption that the interneuron circuit performs compartment-specific feedback inhibition was correct, a perfect match to cortex is probably not to be expected. Firstly, the pyramidal cell model we used is clearly a very reduced depiction of a real pyramidal cell. Because the inhibitory circuitry is optimized for the nonlinear processing performed by these cells, anything that is wrong in the pyramidal cell model will also be wrong in the optimized circuit. It will be interesting to see how the suggested optimization framework generalizes to computations performed by more complex neuronal morphologies (Poirazi and Papoutsi, 2020). A key challenge in this regard will be the choice of an appropriate computational objective. The objective of E/I balance across different compartments may not generalize to more complex morphologies because it is unclear if pyramidal cells can multiplex across more than two compartments – because that would require more different spike patterns – let alone that interneuron circuits could invert such a neural code. Secondly, the optimized circuitry is also sensitive to other modelling choices. For example, the circuit separates spikes and bursts by a synergy between short-term plasticity and interneuron connectivity. A wrong short-term plasticity model will therefore lead to a wrong connectivity in the circuit. Here, it will be interesting to see how a more expressive model of short-term plasticity (Rossbroich et al., 2021)

influences the optimal circuit structure. Finally, of course, our optimality assumption could be wrong to different degrees. We could be wrong in detail: Even if the idea of compartment-specific feedback inhibition were correct, our mathematical representation thereof – matching excitation and inhibition in time – could be wrong, with corresponding repercussions in the optimized circuit. Or we could be wrong altogether: PV and SST interneurons serve an altogether different function, and feedback inhibition is merely a means to a completely different end, such as behavioural circuit modulation (Gentet et al., 2012; Muñoz et al., 2017) or the control of plasticity (Letzkus et al., 2015; Adler et al., 2019). Here, we did not consider alternative functions for the PV-SST diversity, but this would be an interesting topic for future work.

Conclusion

Notwithstanding the dependence of the final circuit on specific model choices, we believe that the suggested optimization approach provides a broadly applicable schema for analyses of structure-function relations of interneuron circuits. On a coarser level of biological detail, optimization approaches have recently been quite successful at linking abstract computations to the neural network level (Mante et al., 2013; Sohn et al., 2019; Mastrogiuseppe and Ostojic, 2018). While similar in spirit, our approach takes this optimization ansatz from the level of dynamical systems analyses of rate-based recurrent neural networks to the detailed level of spiking circuits with multi-compartment neurons and short-term plasticity. It will be exciting to see how biological mechanisms on this level of detail support more advanced computations than the mere stabilization of the circuit considered here, but that is clearly a larger research program that extends well beyond the proof of concept presented here.

4.5 Methods

Code and trained models will be made publicly available upon publication, via https://github.com/sprekelerlab/optimizing-interneuron-circuits. We used Python (Van Rossum and Drake, 2009) version 3.7.3, Numpy (Harris et al., 2020) version 1.18.5, PyTorch (Paszke et al., 2019a) version 1.5.1, and scikit-learn (Pedregosa et al., 2011) version 0.23.2.

Network Model

We simulated a spiking network model consisting of N_E pyramidal cells (PCs) and N_I interneurons (INs), as in earlier work (Naud and Sprekeler, 2018). PCs are described by a two-compartment model (Naud et al., 2014). The membrane potential v^s in the somatic compartment is modelled as a leaky integrate-and-fire unit with spike-triggered adaptation:

$$\frac{dv^s}{dt} = -\frac{v^s - E_L}{\tau_s} + \frac{g_s f(v^d) + w^s + I^s}{C_s} \tag{4.2}$$

$$\frac{dw^s}{dt} = -\frac{w^s}{\tau_{s,w}} + b_s S(t)\,. \tag{4.3}$$

Here, E_L denotes the resting potential, τ_s the membrane time constant and C_s the capacitance of the soma. I^s is the external input, and w^s the adaptation variable, which follows leaky

dynamics with time constant $\tau_{s,w}$, driven by the spike train S emitted by the soma. b_s controls the strength of the spike-triggered adaptation. v^d is the dendritic membrane potential, the conductance g_s controls how strongly the dendrite drives the soma, and f the nonlinear activation of the dendrite:

$$f(v) = 1/(1 + \exp(-(v - E_d)/D_d)) \,. \tag{4.4}$$

The half-point E_d and slope D of the transfer function f control the excitability of the dendrite. When the membrane potential reaches the spiking threshold ϑ, it is reset to the resting potential, and the PC emits a spike. Every spike is followed by an absolute refractory period of τ_r.

The dynamics of the dendritic compartment are given by:

$$\frac{dv^d}{dt} = -\frac{v^d - E_L}{\tau_d} + \frac{g_d f(v^d) + c_d K(t - \hat{t}) + w^d + I^d}{C_d} \tag{4.5}$$

$$\frac{dw^d}{dt} = -\frac{w^d}{\tau_{d,w}} + \frac{a_d(v^d - E_L)}{\tau_{d,w}}. \tag{4.6}$$

In addition to leaky membrane potential dynamics with time constant τ_d, the dendrite shows a voltage-dependent nonlinear activation f, the strength of which is controlled by g_d. This nonlinearity allows the generation of dendritic plateau potentials ("calcium spikes"). Somatic spikes trigger backpropagating action potentials in the dendrite, modelled in the form of a boxcar kernel K, which starts 1 ms after the spike and lasts 2 ms. The amplitude of the backpropagating action potential is controlled by the parameter c_d. The dendrite is subject to a voltage-activated adaptation current w^d, which limits the duration of the plateau potential. This adaptation follows leaky dynamics with time constant $\tau_{d,w}$. The strength of the adaptation is given by the parameter a_d. Note that the model excludes sub-threshold coupling from the soma to the dendrite.

The interneurons are modeled as leaky integrate-and-fire neurons:

$$\frac{dv^i}{dt} = -\frac{v^i - E_L}{\tau_i} + \frac{I^i}{C_i} \,, \tag{4.7}$$

with time constant τ_i. Spike threshold, resting and reset potential, and refractory period are the same as for the PCs.

All neurons receive an external background current to ensure uncorrelated activity, which follows Ornstein-Uhlenbeck dynamics

$$\frac{dI^{x,bg}}{dt} = -\frac{I^{x,bg} - \mu_x}{\tau_{bg}} + \sigma_x \varepsilon. \tag{4.8}$$

Here, $x \in \{s, d, i\}$ refers to the soma, dendrite, or interneuron, respectively, and ε is standard Gaussian white noise with zero mean and correlation $\langle \varepsilon(t)\varepsilon(t') \rangle = \delta(t - t')$.

In addition, the somatic and dendritic compartments received step currents mimicking external signals (see Optimization), as well as recurrent inhibitory inputs. The recurrent input to compartment $x \in \{s, d\}$ of the ith principal cell was given by

$$I_i^{x,inh}(t) = -\sum_{j=1}^{N_I} |W_{ij}^{I\to x}|\ s^j(t). \tag{4.9}$$

where s^j is the synaptic trace that is increased at each presynaptic spike and decays with time constant τ_{syn} otherwise:

$$\frac{ds}{dt} = -\frac{s}{\tau_{syn}} + S.$$

The compartment-specific inhibitory weight matrices $W^{I\to x}, x \in \{s, d\}$ were optimized; the absolute value in Eq. 4.9 ensured positive weights.

The recurrent input to the ith interneuron was given by:

$$I_i^{rec} = \sum_{j=1}^{N_E} |W_{ij}^{E\to I}|\ \mu_{ij}(t)\ s^j(t) - \sum_{k=1} |W_{ik}^{I\to I}| s^k(t). \tag{4.10}$$

The function $\mu_{ij}(t)$ implements short-term plasticity according to the Tsodyks-Markram model (Tsodyks et al., 1998). $\mu(t)$ is the product of a utilization variable u and a recovery variable R that obey the dynamics

$$\frac{du}{dt} = -\frac{u - U}{\tau_u} + (1 - u) \cdot F \cdot S, \tag{4.11}$$

$$\frac{dR}{dt} = -\frac{R - 1}{\tau_R} - u \cdot R \cdot S. \tag{4.12}$$

U is the initial release probability, which is optimized by gradient descent. F is the facilitation fraction, and τ_R, τ_u are the time constants of facilitation and depression, respectively. All parameter values are listed in Table A.1 (Supplementary Materials)

Finally, the network parameters were scaled so that the membrane voltages ranged between $E_L = 0$ and $\vartheta = 1$. The scaling allowed weights of order $1/\sqrt{N}$, mitigating vanishing or exploding gradients during optimization. All optimization parameters are listed in Table A.2 (Supplementary Materials).

Optimization

We used gradient descent to find weights W and initial release probabilities U that minimize the difference between excitation and inhibition in both compartments:

$$\mathcal{L} = \sum_{t=1}^{T} \sum_{i=1}^{N_E} \left(E_i^s(t) + I_i^s(t)\right)^2 + \left(E_i^d(t) + I_i^d(t)\right)^2. \tag{4.13}$$

E_i^x and I_i^x are the total excitatory and inhibitory input to compartment $x \in \{s, d\}$ of PC i. In most simulations (all except those for figure A.4), the output synapses from a given neuron to

a given compartment type had the same strength, i.e., the optimization of the output synapses is performed for $N_I \times 2$ parameters. For the input synapses onto the INs, weight and initial release probability were optimized independently for all $N_E \times N_I$ synapses.

To achieve small interneuron rates necessary for interneuron specialization (Fig. A.2), we subtracted the mean background input from E_i^x:

$$I_i^x(t) = E_i^x(t) - \mu_x, \tag{4.14}$$

such that the interneurons did not fire when their target compartment received its minimum level of external excitation. To propagate gradients through the spiking non-linearity, we replaced its derivative with the derivative of a smooth approximation (Zenke and Ganguli, 2018)

$$\sigma(v) = \frac{1}{(1 + \beta|v - \vartheta|)^2}. \tag{4.15}$$

We used (surrogate) gradient descent instead of gradient-free methods because of its favourable sample efficiency and its recent success in optimizing large-scale spiking networks (Bellec et al., 2018; Zenke and Vogels, 2021). We used the machine learning framework PyTorch (Paszke et al., 2019b) to simulate the differential equations (forward Euler with step size 1 ms), compute the gradients of the objective $\mathcal{L}$ using automatic differentiation, and update the network parameters using Adam (Kingma and Ba, 2014). Backpropagation through time requires storing intermediate activation values during the forward pass (network simulation), followed by a backward pass. Our network consists not just of multi-compartmental neurons but also of a short-term plasticity model that introduces N^2 additional variables (N being the network size). Scaling the model to larger network sizes might therefore require approximating gradient descent by a local learning rule or by gradient-free optimization. The optimized parameters were initialized according to the distributions listed in Table A.2 (Supplementary Materials). A.2. We simulated the network response to batches of 8 trials of 600 ms, consisting of 100 ms pulses given at 2.5 Hz. The pulse amplitudes were drawn uniformly and independently for soma and dendrites from the set $\{100, 200, 300, 400\}$. Training converged within 200 batches (parameter updates). Before each parameter update, the gradient values were clipped between -1 and 1 to mitigate exploding gradients (Pascanu et al., 2013). After each update, the initial release probability was clipped between 0 and 1 to avoid unphysiological values. We trained networks without clipping the gradient or the release probability to confirm that this did not bias the solutions found by the optimization.

Methods for Figures

Figure 4.1

We measured the short-term plasticity of PC $\rightarrow$ IN synapses by simulating their response to two EPSPs given 10 ms apart, a typical interspike interval within a burst. The paired pulse ratio (PPR) was computed as the ratio of the two excitatory postsynaptic potential (EPSP) amplitudes, such that a PPR > 1 indicates short-term facilitation and a PPR < 1

indicates short-term depression. The PPR of a single IN was defined as the mean PPR of all its excitatory afferents. Clustering of interneurons was done by fitting a single Gaussian (before optimization) or a mixture of two Gaussians (after optimization) to the three-dimensional distribution of inhibitory weights to the PC soma, to PC dendrites, and the PC→IN Paired Pulse Ratio (PPR). Both models were fitted using Scikit-learn (Pedregosa et al., 2011) on pooled data from five networks trained from different random initializations. The density models were fitted on 246 interneurons that were active (firing rate higher than 1 spk/s) and had a medium to strong projection to either soma or dendrites (weight bigger than 0.01). The dashed lines in Fig. 4.1b illustrate the two-dimensional marginal distributions of the somatic and dendritic inhibition. All PCs received the same time-varying input currents, consisting of 100 ms pulses of 300 pA, given at a rate of 2.5 Hz. Correlations between compartment-specific excitation and inhibition were computed between the currents to the PC compartments, averaged across all PCs in the network.

Figure 4.2

Before optimization, we assigned interneurons to inhibit either PC somata or dendrites by fixing their weights onto the other compartment to zero. Half of the interneurons were assigned to inhibit the soma; the other half was assigned to inhibit the dendrites. Otherwise, weights and initial release probabilities were optimized as before.

Figure 4.3

The definitions of burst rate, burst probability, and event rate were taken from Naud & Sprekeler (Naud and Sprekeler, 2018): A burst was defined as multiple spikes occurring within 16 ms. The time of the first spike was taken as the time of the burst. An event was defined as a burst or a single spike. The instantaneous burst rate and event rate were computed by counting the number of bursts and events, respectively, in bins of 1ms and among the population of PCs and smoothing the result with a Gaussian filter (width: 2ms). The burst probability was defined as

$$\text{Burst Probability} = \frac{\text{Burst Rate}}{\text{Event Rate}} \times 100\%. \tag{4.16}$$

We injected current pulses of 100 ms duration to either soma or dendrite while injecting a constant current to the other compartment. Currents were varied in amplitude between 100 and 400 pA; the constant current was 0 pA. The figure shows the mean and standard deviation of the total network activity during 10 current pulses. For Fig. 4.3e, we injected simultaneous pulses to the other compartment of amplitude $0, 200$ or 400 pA.

Figure 4.4

We varied the correlation between the inputs to soma and dendrites by generating repeating current pulses with different temporal offsets and optimized a network for each offset. The

interneuron specialization was defined as

$$\text{specialization} = 1 - \frac{x^T y}{\|x\|\|y\|}\,, \tag{4.17}$$

where x and y are N_I-dimensional vectors containing the inhibitory weights onto soma and dendrites and $\|\cdot\|$ the L_2 norm. If each neuron inhibits either somata or dendrites but not both, the specialization will be 1. If the weights are perfectly aligned (i.e., interneurons with a strong dendritic projection also have a strong somatic projection), the specialization will be 0. Here and in all figures, the E/I correlation was computed as the correlation between the time series of the compartment-specific excitation and inhibition after averaging across all PCs. Shown is the mean over 5 batches of 600 ms, where each batch consisted of 8 trials with amplitudes from $\{100, 200, 300, 400\}$ pA, sampled independently for soma and dendrites.

Figure 4.5

Figure 4.5a shows the connectivity strength over five networks. We first used the Gaussian mixture models to assign INs to PV or SST clusters, and then computed the mean connectivity between and within clusters for each network. For 4.5b, we trained networks with predefined interneuron populations to control the interneuron connectivity. Connections between populations were knocked out by fixing them to zero during and after optimization. E/I correlations are computed for 5 batches of 600 ms, where each batch consisted of 8 trials with amplitudes from $\{100, 200, 300, 400\}$ pA, sampled independently for soma and dendrites.

Figure A.1

As for Fig. 4.2, we assigned interneurons to inhibit either PC somata or dendrites. Here, we trained networks for different correlations between compartment-specific external inputs (cf. Fig. 4.4) and baseline activity levels (Supplementary Materials, Fig. A2). We used the 10th percentile as a robust measure of minimum PV rate. The mean and sd PPR of the PV and SST populations computed over all INs that were active (rate larger than 1 spk/s) and provided a medium to strong inhibition to one PC compartment (weight bigger than 0.01).

Figure A.2

The minimum rate of PV neurons was controlled indirectly by varying the baseline inhibitory target current to the soma—A larger baseline requires a higher minimum PV rate. We varied the minimum inhibitory current by subtracting only a fraction α of the baseline excitatory current:

$$I^x(t) = E^x(t) - \alpha \cdot \mu_x, \tag{4.18}$$

cf. Eq. (4.14). In the simulations, we varied α between 1 and 0.8, leading to a minimum PV rate between 1 spk/s, and 9 spk/s.

Figure 4.4

We trained networks without short-term plasticity by setting the learning rate of the initial release probability U to 0. We ran the optimization for 1600 instead of the usual 400 steps to ensure that differences between network performance with and without short-term plasticity were not due to slower convergence. We also trained networks using a range of learning rates (between $[0.0001, 0.01]$) to ensure any difference could not be due to a sub-optimal learning rate. The optimal learning rate for these non-STP simulations was the same as the default rate for the STP simulations. We trained 5 networks starting from different random weight initialization, and having confirmed that the results were similar across networks, we picked one at random to generate the figure.

Figure A.4

This figure shows networks in which interneurons project to PC somata and dendrites with a PC-specific weight. That is, the inhibition to PC somata is given by a $N_I \times N_E$ matrix $W^{I \to S}$, and the inhibition to PC dendrites is given by a matrix $W^{I \to D}$ of the same dimensions. In the other simulations, both inhibitory weights were defined by $N_I \times 1$ matrices, such that a particular interneuron projected to all PC somata with the same weight and to all PC dendrites with the same weight. Clustering of interneurons was done in the two-dimensional weight space, defined by their mean inhibition to soma and mean inhibition to dendrites (averaged over all PCs).

5

Cortical interneurons: fit for function and fit to function? Evidence from development and evolution

5.1 Context within thesis

The previous chapter showed that optimising interneurons for a specific function leads to the emergence of cells resembling PV+ and SST+ interneurons. This function therefore explains the existence of these cell types in our model. Does it also explain the existence of PV+ and SST+ cells in the brain? To find out, we revisited data on the biological emergence of these cell types. Specifically, we asked whether interneuron development or evolution could be driven by a functional process similar to our model's optimization. Crucially, this would require the development or evolution of pyramidal cells to precede that of interneurons.

Our findings suggest the opposite: interneurons and their specific features (short-term plasticity and, possibly, connectivity) likely predate pyramidal cells and their specific features (dendrite-dependent bursting). This is true during development as well as evolution. The model from Ch. 4 therefore captures a possible function of interneurons in the mature mammalian brain but does not explain how these interneurons came to be.

The dissociation between theoretical function and evolutionary history highlights the need to empirically test adaptive hypotheses (Ch. 7).

5.2 Abstract

Cortical inhibitory interneurons form a broad spectrum of subtypes. This diversity suggests a division of labour in which each cell type supports a distinct function. In the present era of optimisation-based algorithms, it is tempting to speculate that these functions were the evolutionary or developmental driving force for the spectrum of interneurons we see in the mature mammalian brain. In this study, we evaluated this hypothesis using the two most

common interneuron types, parvalbumin (PV) and somatostatin (SST) expressing cells, as examples. PV and SST interneurons control the activity in the cell bodies and the apical dendrites of excitatory pyramidal cells, respectively, due to a combination of anatomical and synaptic properties. But was this compartment-specific inhibition indeed the function for which PV and SST cells originally evolved? Does the compartmental structure of pyramidal cells shape the diversification of PV and SST interneurons over development? To address these questions, we reviewed and reanalysed publicly available data on the development and evolution of PV and SST interneurons on one hand and pyramidal cell morphology on the other. These data speak against the idea that the compartment structure of pyramidal cells drove the diversification into PV and SST interneurons. In particular, pyramidal cells mature late, while interneurons are likely committed to a particular fate (PV vs. SST) during early development. Moreover, comparative anatomy and single-cell RNA-sequencing data indicate that PV and SST cells, but not the compartment structure of pyramidal cells, existed in the last common ancestor of mammals and reptiles. Specifically, turtle and songbird SST cells also express the *Elfn1* and *Cbln4* genes that are thought to play a role in compartment-specific inhibition in mammals. PV and SST cells therefore evolved and developed the properties that allow them to provide compartment-specific inhibition before there was selective pressure for this function. This suggests that interneuron diversity originally resulted from a different evolutionary driving force and was only later co-opted for the compartment-specific inhibition it seems to serve in mammals today. Future experiments could further test this idea using our computational reconstruction of ancestral Elfn1 protein sequences.

5.3 Introduction

Cortical inhibitory interneurons are a highly diverse group, differing in their morphology, connectivity, and electrophysiology (Tremblay et al., 2016). Decades of experimental and theoretical work have suggested a role for interneurons in many functions (Tremblay et al., 2016; Kepecs and Fishell, 2014; Sadeh and Clopath, 2021), including the regulation of neural activity (Vogels et al., 2011; Wu et al., 2022), control of synaptic plasticity (Letzkus et al., 2015; Williams and Holtmaat, 2019), increasing temporal precision (Wehr and Zador, 2003; Bhatia et al., 2019), predictive coding (Keller and Mrsic-Flogel, 2018; Hertäg and Clopath, 2022), and gain modulation (Fu et al., 2014; Ferguson and Cardin, 2020). Many of these functions come down to the control of excitation.

Why would the control of excitation require a diversity of interneurons? A key reason could lie in the complexity of excitatory cells (Fishell and Kepecs, 2020; Keijser and Sprekeler, 2022). Pyramidal cells (PCs) consist of several cellular compartments that have different physiological properties (e.g. sodium vs. calcium spikes (Larkum et al., 1999)), receive different inputs (e.g., top-down vs. bottom-up (Petreanu et al., 2007; Larkum, 2013), although see (Ledderose et al., 2022)) and follow distinct synaptic plasticity rules (Letzkus et al., 2006; Sjostrom et al., 2008; Udakis et al., 2020). The control of different pyramidal cell compartments might therefore require inhibition from designated types of interneurons. Indeed, the two most common interneuron types—parvalbumin (PV)- and somatostatin (SST)-expressing cells—are classically distinguished by their connectivity with pyramidal cells: whereas PV-expressing

basket cells mainly target the somata of PCs, SST-expressing Martinotti cells mainly target their apical dendrites (Tremblay et al., 2016). The cellular and synaptic properties of these interneurons also seem adapted to this purpose. SST interneurons receive facilitating synapses from PCs (Reyes et al., 1998; Silberberg and Markram, 2007), rendering them sensitive to bursts of action potentials (Berger et al., 2010; Murayama et al., 2009; Goldberg et al., 2004) triggered by plateau potentials in the apical dendrite of PCs (Larkum et al., 1999; Williams and Stuart, 1999). Indeed, SST interneurons control dendritic excitability and bursting of PCs (Murayama et al., 2009; Gentet et al., 2012; Lovett-Barron et al., 2012). PV interneurons, on the other hand, receive depressing synapses (Reyes et al., 1998; Caillard et al., 2000), rendering them less sensitive to these signals (Pouille and Scanziani, 2004). The presynaptic dynamics of PV and SST interneurons therefore seem particularly well-matched to the physiology of pyramidal cells, although both types also inhibit non-pyramidal cells and other interneurons (see e.g. Jiang et al. (2015) and Campagnola et al. (2022)). These and similar observations have led to the view that interneuron diversity can be understood from a functional perspective, in which the morphology and synaptic and cellular properties of different interneurons are fit to specific functions (Fig. 5.1a; Kepecs and Fishell (2014) and Fishell and Kepecs (2020)). Consistent with this idea that interneurons are adapted to control different pyramidal cell compartments, we recently showed that properties (connectivity and short-term plasticity) of PV and SST interneurons emerge when optimising a network model for compartment-specific inhibition (Fig. 5.1b; Keijser and Sprekeler (2022)).

The specialisation of PV and SST interneurons to pyramidal soma and dendrites, respectively, makes it tempting to speculate that the diversification of these interneuron subtypes was driven by pyramidal cell properties, either during evolution or during development (Fig. 5.1c). This hypothesis predicts a specific temporal order: During evolution or development, the compartmentalization of pyramidal cells should predate interneuron diversification (Fig. 5.1d).

Here, we evaluate this idea with a focus on PC and interneuron properties that seem particularly well-adapted to each other: the active dendrites of pyramidal cells and the connectivity and short-term plasticity of interneurons. Reviewing and re-analysing recent evolutionary and developmental data, we reconstruct the developmental and evolutionary history of these three properties. We find no support for the idea that interneurons develop or evolved to control preexisting compartments of pyramidal cells. Instead, the central properties of PV and SST interneurons that led to this idea emerged before the PC properties they seem adapted to in both development and evolution. Rather than pyramidal physiology driving interneuron diversification, this suggests a model in which existing interneuron properties enabled new pyramidal cell functions.

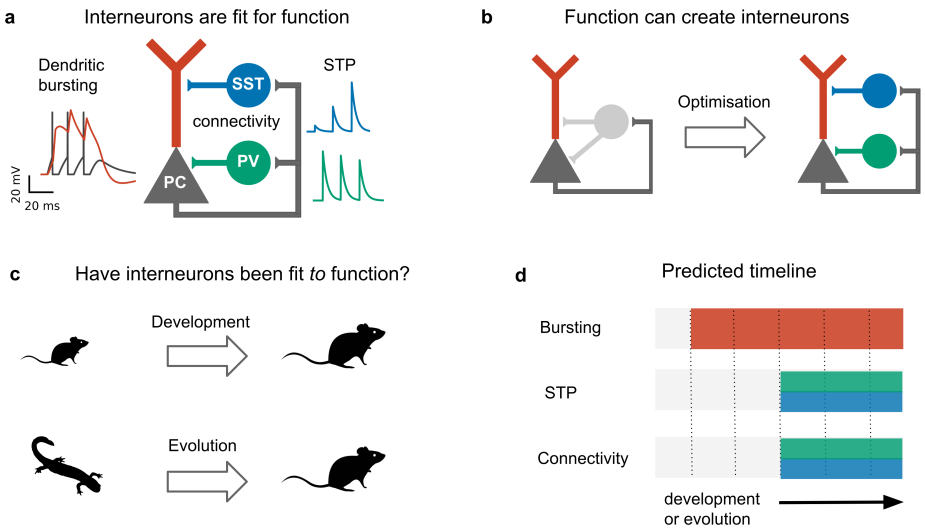

Figure 5.1: Do development and evolution fit interneurons to function? **(a)** The connectivity and short-term plasticity (STP) of PV and SST positive interneurons seem adapted to the morphology and electrophysiology of pyramidal cells, as highlighted by an optimization-based model **(b)**. In this model, optimising interneuron parameters to provide compartment-(soma/dendrite) specific inhibition causes interneurons to diversify into two groups that resembled PV and SST interneurons in their connectivity and short-term plasticity (Keijser and Sprekeler, 2022). **(c)** The existence of PV and SST subtypes might therefore result from a developmental or evolutionary tuning of interneurons based on pyramidal properties. **(d)** This predicts that immature (or ancestral) circuits contain bursting pyramidal neurons and undiversified interneurons. Development (or evolution) then diversifies the interneurons into PV and SST subtypes. PC activity and short-term plasticity simulated with models from refs. (Naud et al., 2013; Naud and Sprekeler, 2018) and (Tsodyks et al., 1998), respectively. Animal silhouettes from `https://beta.phylopic.org/`

5.4 Developmental trajectory of compartment-specific inhibition

We first discuss the developmental trajectory of pyramidal cells and PV and SST interneurons in the mammalian cortex to assess whether the diversification of PV and SST interneurons during development is driven by pyramidal cell properties. We mostly consider data from rodents, but many of the findings seem to be conserved among mammals (Hansen et al., 2013; Ma et al., 2013; Shi et al., 2021; Schmitz et al., 2022).

In contrast to pyramidal cells, interneurons are not born in the developing cortex but subcortically (Anderson et al., 1997; Flames et al., 2007). It is only upon migrating to the cortex that interneurons acquire their mature morphology and physiology. The long period between interneuron birth and maturation has led to different models of interneuron development (Kepecs and Fishell, 2014; Wamsley and Fishell, 2017). One model attributes the late maturation of interneurons to a late specification of their cellular identity, possibly based on external cues within the circuit they embed themselves in (Kepecs and Fishell, 2014; Wamsley and Fishell, 2017). Alternatively, the late emergence of characteristic features could be due to the slow unfolding of a predetermined genetic program that happens independently

of the surrounding circuit(Lim et al., 2018a).

The malleability of interneuron properties during development is therefore currently an open question: Which properties are adapted to the surrounding circuit, and which are predetermined? Whatever properties are adapted, cellular identity (e.g., PV vs SST) is probably not one of them (Wamsley and Fishell, 2017; Lim et al., 2018a). Interneuron types—at least on a coarse level—are determined by their time and place of birth. Future PV and SST interneurons, for example, are preferentially generated within different parts of the same embryonic structure (Wonders and Anderson, 2006; Lim et al., 2018a).

Recent data suggests that not just interneuron types (e.g., PV vs. SST) but also interneuron *sub*types (e.g., SST Martinotti vs. SST non-Martinotti) are specified early in development. Lim et al. (Lim et al., 2018a) showed that Martinotti and non-Martinotti cells migrate to the developing cortex via different routes. In addition, a developing interneuron's transcriptional profile can be used to predict its future fate (Mi et al., 2018; Mayer et al., 2018; Shi et al., 2021; Bandler et al., 2021).

Although interneurons are, therefore, likely hardwired to become a certain subtype, it is still possible that interneuron properties such as short-term plasticity or connectivity are subject to activity-dependent fine-tuning. For example, the development of short-term facilitation or a layer 1 axon of SST Martinotti cells might emerge in dependence on pyramidal neuron bursting. In this case, bursting should develop ahead of these SST features.

Birth and migration of cortical interneurons

Cortical GABAergic interneurons are born in a transient region of the developing brain known as the ganglionic eminence (Anderson et al., 1997; Lim et al., 2018a; Wamsley and Fishell, 2017), from where they tangentially migrate to the cortex (Marín and Rubenstein, 2001). The ganglionic eminence can be divided into multiple subregions patterned by unique combinations of transcription factors (Flames et al., 2007; Hansen et al., 2013; Ma et al., 2013; Hu et al., 2017) that activate distinct genetic programs. Since each genetic program corresponds to a different cell type, the majority of the cells born in the medial ganglionic eminence (MGE) will become PV and SST interneurons, whereas the caudal ganglionic eminence (CGE) generates, among others, vasoactive intestinal peptide (VIP)-expressing interneurons (Wichterle et al., 2001; Nery et al., 2002; Xu et al., 2004; Butt et al., 2005; Miyoshi et al., 2010). A key example of a patterning transcription factor that shapes interneuron identity is *Nkx2-1*, which is expressed within the MGE but not CGE (Sussel et al., 1999; Butt et al., 2005). *Nkx2-1* knockout leads MGE-derived interneurons to adopt the fate of CGE-derived interneurons (Butt et al., 2008). Molecular gradients have also been shown to contribute to interneuron diversity within the same eminence: the dorsal-caudal and rostral-ventral MGE preferentially generate SST and PV neurons, respectively (Wonders et al., 2008; Fogarty et al., 2007; Inan et al., 2012; He et al., 2016; Hu et al., 2017; McKenzie et al., 2019).

After birth, interneurons migrate to the developing cortex via two different routes: The superficial marginal zone (the MZ, which will develop into cortical layer 1) and the deeper subventricular zone (SVZ). These different migration routes are used by distinct layer 2-3 (L2-3) SST subtypes (Lim et al., 2018a). Whereas L2-3 SST Martinotti cells migrate via the marginal zone from where they descend to their future location, non-Martinotti cells migrate via the subventricular zone (Lim et al., 2018b). This indicates that SST subtypes (at least in L2-3) are predetermined before their arrival in cortex, possibly even before they "choose" one migratory route over the other. Future L2-3 Martinotti cells forced to migrate via the wrong route (the SVZ) still become Martinotti cells in terms of their transcriptional profile and electrophysiology, but they lack a fully developed layer 1 axon (Lim et al., 2018b). This suggests that developing L2/3 Martinotti cells cannot send their developing axon from deeper to upper layers but must leave it there while their cell body descends. Translaminar axons of other neurons, such as a less studied PV subtype (Lim et al., 2018b), and cerebellar granule cells (Rakic, 1971) are established via a similar mechanism, suggesting it might be the only reliable way for neurons to develop translaminar projections.

When do developing pyramidal cells first show dendrite-dependent bursting? Their electrophysiology matures relatively late: dendritic plateau potentials emerge only in the third postnatal week (Zhu, 2000; Franceschetti et al., 1998). This is consistent with the late maturation of their dendritic morphology. PCs develop their intricate apical arborization

and tuft dendrites after the second postnatal week (Zhu, 2000; Romand et al., 2011). For example, the tuft length increases almost twofold during the third postnatal week (Romand et al., 2011), and dendritic spikes fail to reach the soma on postnatal day 14 and 28 (Zhu, 2000).

When does short-term facilitation (STF) of PC→SST synapses arise during development? Could its development be driven by bursting in pyramidal cells? Some of the early experiments showed such STF in rat cortex during the third postnatal week (Beierlein and Connors, 2002; Silberberg and Markram, 2007; Reyes et al., 1998). Evidence for an even earlier presence of STF in these synapses comes from molecular studies. The Gosh laboratory has shown that the short-term facilitation in hippocampal (Sylwestrak and Ghosh, 2012) and cortical (Stachniak et al., 2019) PC→SST synapses is due to the transmembrane protein Elfn1, which is expressed by SST mouse and human interneurons (Fig. 5.2). In these experiments, STF was measured in the second postnatal week, and the expression of Elfn1 was detected already one week after birth (Tomioka et al., 2014; Favuzzi et al., 2019), providing an early molecular signature of short-term facilitation in SST cells. Short-term facilitation in PC→SST synapses is therefore present before dendrite-dependent bursting in PCs.

Genetic basis of short-term facilitation

Pyramidal cells form short-term depressing synapses onto PV neurons but short-term facilitating synapses onto SST neurons. This difference is partly attributed to the postsynaptic expression of Elfn1 by SST neurons (Sylwestrak and Ghosh, 2012; Tomioka et al., 2014; Stachniak et al., 2019). Elfn1 is a synaptic protein that contacts the presynaptic boutons of pyramidal cells and controls their release properties. Specifically, Elfn1 induces presynaptic localization of metabotropic glutamate receptor 7 (mGluR7) (Tomioka et al., 2014). *Grm7*, the gene coding for mGluR7, is near-ubiquitously expressed in mouse (and human) neurons (data from Tasic et al. (2018) and Bakken et al. (2021)). mGluR7 has a low affinity for glutamate: Only high glutamate levels caused by repeated presynaptic stimulation will lead mGluR7 to activate calcium channels, which increase synaptic release and thereby mediate synaptic facilitation. Elfn1 causes facilitation of PC→SST synapses in the hippocampus and different cortical layers (Stachniak et al., 2019). As expected from their expression of Elfn1, human SST (and VIP) interneurons receive facilitating inputs (Campagnola et al. (2022); Fig. 5.2) However, in the mouse brain, the correlation between the short-term facilitation and the expression of Elfn1 is very high, but not perfect (Stachniak et al., 2021).

What about the second difference between PV and SST neurons, their compartment-specific output synapses? SST and PV cells form compartment-specific synapses in visual cortical organotypic cultures that lack external inputs (Cristo et al., 2004). This strongly suggests a role for genetic encoding rather than experience-dependent activity. Indeed, recent work identified important molecular players in establishing compartment-specific synapses (Favuzzi et al., 2019). Several genes are involved in the formation of compartment-specific synapses. For example, suppressing *Cbln4* in SST interneurons decreased inhibition of PC dendrites. An

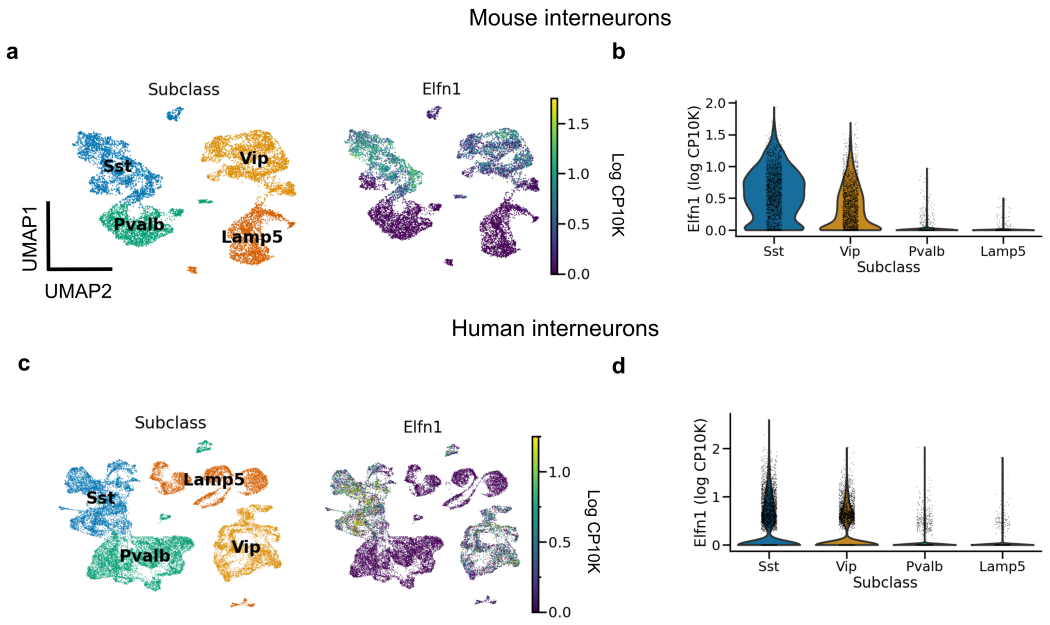

Figure 5.2: *Elfn1* expression correlates with short-term facilitation in mammals. (a) UMAP (McInnes et al., 2018) plot of mouse interneurons coloured by subclass (left) and *Elfn1* expression (right). The two interneuron types — SST and VIP interneurons — known to receive facilitating synapses both express *Elfn1*. **(b)** Violin plot of *Elfn1* expression by subclass. CP10K: counts per 10 thousand. **(c,d)** As **(a,b)**, but for human interneurons. Data from from refs. (Tasic et al., 2018) (a,b), and (Bakken et al., 2021) (c,d).

over-expression of the same gene in PV interneurons, on the other hand, increased inhibition of PC dendrites (Favuzzi et al., 2019). Other genes contribute to somatic inhibition in a seemingly analogous way (Favuzzi et al., 2019). Both loss and gain of function were shown around P14. Similarly, somatic inhibition in CA1 abruptly emerges at the end of the second postnatal week (Dard et al., 2022). It is therefore by the second postnatal week that PV and SST interneurons are committed as to where to direct their output synapses.
Intriguingly, *Cbln4* is only expressed in a subset of neurons (Fig. 5.3). Clustering revealed that these *Cbln4*+ neurons correspond to previously identified subtypes. The *Tac1* cluster labels non-Martinotti cells that target the dendrites of L4 cells (Gouwens et al., 2020; Nigro et al., 2018; Scala et al., 2019), and the *Calb2* and *Etv1* clusters correspond to fanning-out Martinotti cells (Gouwens et al., 2020; Wu et al., 2022), consistent with a role of *Cbln4* in establishing dendritic synapses. But only a subset of the *Myh8* cluster—corresponding to T-shaped Martinotti cells (Wu et al., 2022)— expressed *Cbln4*, suggesting diverse mechanisms for dendritic targeting.

An interneuron's cell type, the plasticity of their input synapses from PCs, and the PC compartments they target are therefore determined before interneurons are fully embedded within cortical circuits and before pyramidal neurons develop their characteristic morphology and electrophysiology. This suggests that while PV and SST interneurons are fit for the function of compartment-specific inhibition of PCs, some of their characteristic properties are probably not developmentally driven by PC activity.

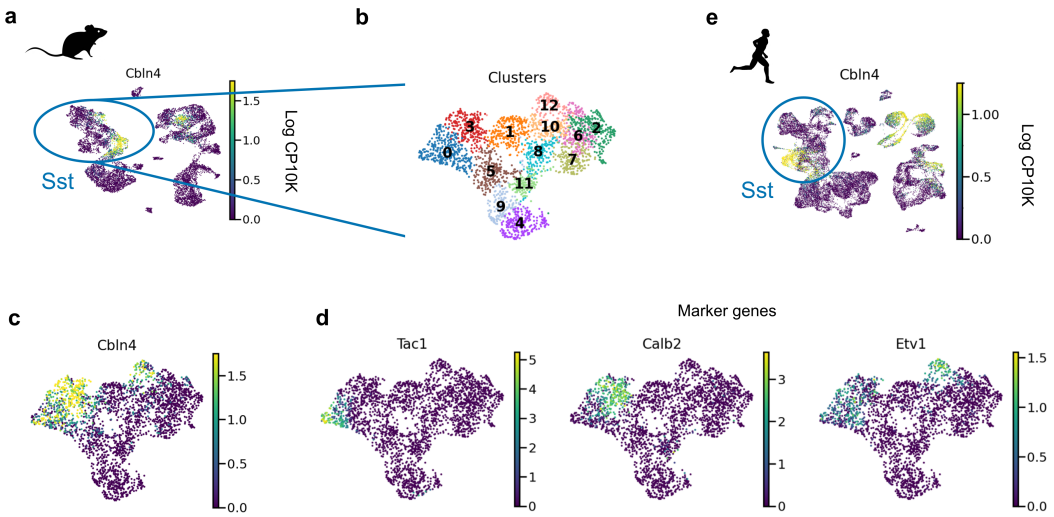

Figure 5.3: *Cbln4* is expressed in a subset of mammalian SST interneurons. (a) UMAP plot of mouse and human interneurons, coloured by their expression of *Cbln4*, a gene that instructs synapse formation onto pyramidal dendrites in mice (Favuzzi et al., 2019). *Cbln4* is expressed in certain mouse and human interneuron subtypes, including SST cells. **(b)** UMAP of SST cells, clustered into subgroups. **(c)** *Cbln4* is expressed in clusters 0, 3, and 12, which also express marker genes *Tac1*, *Calb2* and *Etv1* **(d)**, respectively. **(e)** A subset of human Sst cells also express *Cbln4*. Data from from refs. (Tasic et al., 2018) (a) and (Bakken et al., 2021) (e).

5.5 Evolutionary trajectory of compartment-specific inhibition

On a much longer timescale than development, evolution also changes the properties of cell types. This raises the question of whether the differentiation of PV and SST interneurons preceded the evolution of the compartmental complexity of pyramidal neurons.

If natural selection tuned PV and SST neurons to pyramidal cell properties, the brains of mammalian ancestors must have contained pyramidal cells with elaborate dendrites, while interneurons were still undifferentiated (Fig. 5.1c,d). This hypothesis cannot be tested directly since our mammalian ancestors are no longer alive, and their fossils provide no information regarding cell types. We therefore have to infer the evolutionary history of cell types by comparing data from modern-day species (Fig. 5.4a, (Arendt et al., 2016; Tosches, 2021b)). Although many cell type-specific properties such as short-term plasticity have not been measured in non-standard model organisms (for recent exceptions, see Gidon et al. (2020), Beaulieu-Laroche et al. (2021), and Campagnola et al. (2022)), transcriptomic correlates can be studied using single-cell RNA sequencing (scRNA-seq, (Tang et al., 2009)), offering a means for defining and comparing cell types across species (Tanay and Sebé-Pedrós, 2021; Tosches, 2021b).

Interneuron conservation & principal neuron divergence

The first applications of scRNA-seq in neuroscience profiled cell types in mice (Zeisel et al., 2015; Tasic et al., 2016). More recently, scRNA-seq was used to classify neuron types also in reptiles (Tosches et al., 2018) and songbirds (Colquitt et al., 2021). The evolutionary relationships of reptiles, birds and mammals suggest that a feature found in all three lineages predates their divergence, while a feature found exclusively in the mammalian lineage is,

in fact, a mammalian invention. This idea enables inferring the evolutionary history of interneurons and pyramidal cells (Fig. 5.4a).

Let us first consider the general evolutionary trajectory of excitatory and inhibitory cell types. Tosches et al. (2018) used scRNA-seq to analyse cells from the turtle and lizard forebrain and compare them with previously published mammalian data (Tasic et al., 2016). They found that reptilian inhibitory neurons cluster into groups that roughly correspond to mammalian interneuron types (Tosches et al., 2018). These results extend earlier findings that found similarities between turtle and mammalian interneurons based on marker genes and morphology (Blanton et al., 1987; Reiner, 1993). Colquitt et al. (2021) recently made analogous observations regarding the similarity of songbird and mouse interneurons (Fig. 5.4b,c). The most parsimonious explanation of this sharing of interneuron types is that similar types already existed in a common ancestor of the three lineages rather than convergent evolution in three lineages. This homology is likely due to shared developmental origins: the inhibitory interneurons of birds and reptiles are born within the conserved ganglionic eminences. The fact that interneurons of different lineages are homologous does not mean they are identical. For example, the correlation between mouse PV and SST cells and the best-matching songbird clusters is 0.37 and 0.31, respectively (Fig. 5.4b). This is higher than the correlation between the best-matching glutamatergic types (0.19, see below), but lower than between some of the different cell types within the same species (mouse PV and SST cells: 0.58). Mammalian and non-mammalian interneurons, while homologous, therefore have likely undergone lineage-specific adaptations.

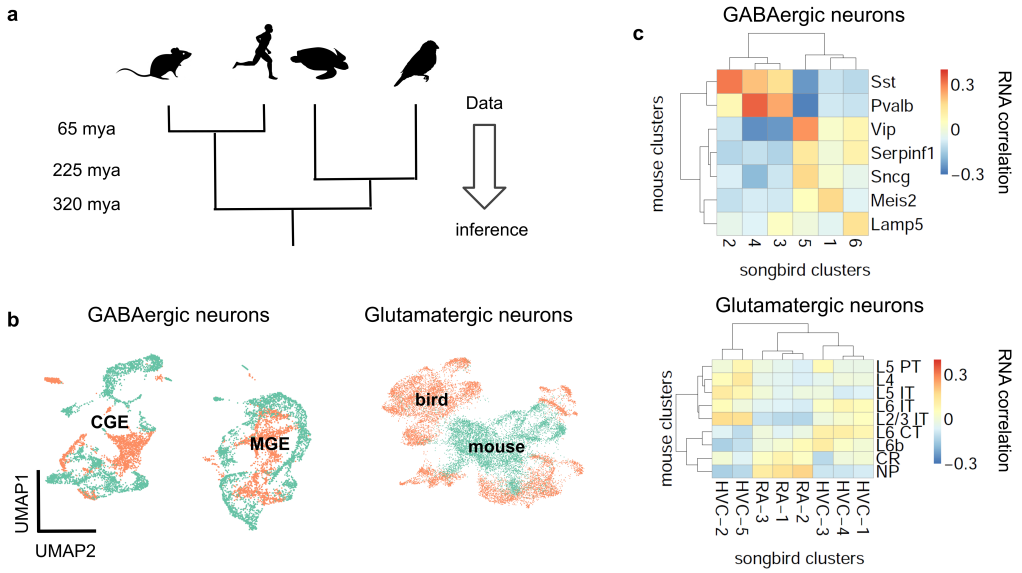

Figure 5.4: Evolutionary conservation of GABAergic cell types. **(a)** Phylogenetic approach. **(b)** Pearson correlation between average RNA expression in clusters of songbird and mouse interneurons. Correlations between GABAergic neurons are typically larger. **(c)** UMAP plots of integrated gene expression data for GABAergic and glutamatergic neurons. GABAergic neurons first cluster by developmental origin (MGE vs. CGE) and then by species. Mouse data from Tasic et al. (2018), songbird data and correlation analysis from Colquitt et al. (2021).

In contrast to inhibitory interneurons, excitatory neurons are probably not homologous between reptiles, songbirds, and mammals (Fig. 5.4b,c; Tosches et al. (2018) and Colquitt

et al. (2021)). Excitatory cell types in different species are defined by different combinations of transcription factors. A clear example is given by the *Fezf2* and *Satb2* genes that specify subcortical (Lodato et al., 2014) and callosal (Alcamo et al., 2008) projections, respectively, of mammalian pyramidal cells. Strikingly, these genes are mutually repressive in the mammalian neurons but co-expressed in reptilian neurons (Tosches et al., 2018; Nomura et al., 2018). Comparing excitatory neurons in the songbird and the mammalian brain revealed an analogous pattern: Although excitatory neurons in the songbird forebrain express similar genes as their counterparts in the mammalian neocortex, these genes are regulated by different transcription factors (Colquitt et al., 2021). Instead, the transcription factors expressed by songbird glutamatergic neurons are similar to those in, e.g. the mouse olfactory bulb and olfactory cortex. Since transcription factors specify cellular identity (Hobert, 2008; Arendt et al., 2016), this suggests that excitatory neurons are not conserved across mammals, birds and reptiles (Tosches et al., 2018; Colquitt et al., 2021; Tosches, 2021a).

Inhibitory cell types therefore seem more conserved than excitatory cell types, which appears broadly inconsistent with an evolutionary adaptation of interneurons to pyramidal cells. This is further confirmed when considering the evolutionary history of specific features of excitatory and inhibitory interneurons, in particular, elaborate dendrites and dendrite-dependent bursting and short-term plasticity.

Evolution of cell type-specific features

We are not aware of direct measurements of short-term facilitation in non-mammalian species and therefore aimed to infer its presence from the expression of *Elfn1*. To this end, we re-analyzed publicly available gene expression data for reptilian and songbird interneuron types (Tosches et al., 2018; Colquitt et al., 2021). We found that *Elfn1* is also expressed in the types corresponding to mammalian SST (and VIP) interneurons (Fig. 5.5) but not in the type corresponding to PV interneurons. This suggests that SST-like interneurons expressing *Elfn1* — and potentially facilitating glutamatergic input synapses — were already present in the last common ancestor of reptiles, songbirds and mammals. In terms of potential transcriptomic correlates of synaptic specificity, we find that *Cbln4* (Fig. 5.3) is expressed in certain subtypes of turtle SST neurons (Fig. 5.6a). Songbird SST neurons, on the other hand, do not express *Cbln4* (Fig. 5.6b). The expression of *Cbln4* by Sst interneurons therefore correlates with the presence of apical dendrites in pyramidal cells (Fig. 5.7, see next). The most parsimonious explanation is that *Cbln4* expression was lost in the songbird lineage. Alternatively, it could have evolved independently in mammals and reptiles.

So not just interneuron subtypes, but also some of their specific properties seem evolutionarily conserved. In contrast, glutamatergic cell types in reptiles and birds show a very different dendritic morphology and physiology from their mammalian pyramidal counterparts. Turtle pyramidal cells have multiple apical dendrites but no basal dendrites (Fig. 5.7,Connors and Kriegstein (1986)). This clear morphological difference suggests that turtle pyramidal neurons are also electrophysiologically distinct. Larkum et al. (2008) showed that *in vitro*, turtle pyramidal neurons lack dendritic calcium spikes and dendrite-dependent bursting.

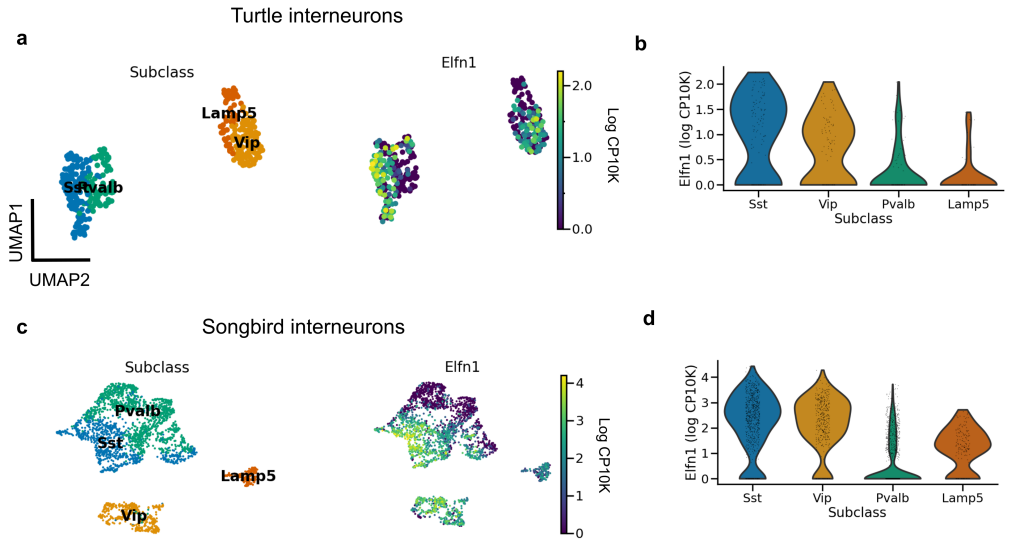

Figure 5.5: Evolutionary conservation of *Elfn1* expression (a) UMAP plot showing over-expression of *Elfn1* in SST-like and VIP-like interneurons in the turtle forebrain. Data from Tosches et al. (2018).**(b)** Violin plots of *Elfn1* expression for each of the clusters. **(c,d)** As (a,b), but for zebra finch neurons. Data from Colquitt et al. (2021).

Morphologically similar pyramidal cells in rodent piriform cortex also lack active dendrites (Bathellier et al. (2009) and Johenning et al. (2009), but see Kumar et al. (2018)). Interestingly, this is probably not due to an absence of calcium channels but rather to the presence of A-type potassium channels (Johenning et al., 2009). Songbird excitatory cells have a stellate morphology and differ, therefore, even more from mammalian pyramidal cells (Fig. 5.7, see, e.g. Devoogd and Nottebohm (1981) and Benezra et al. (2018)).

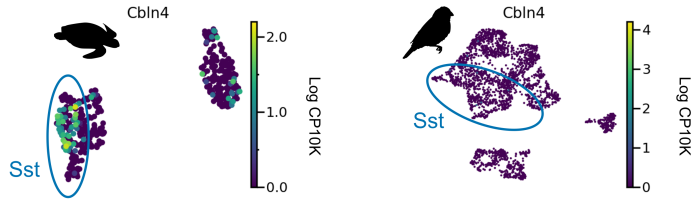

Figure 5.6: *Cbln4* expression in non-mammalian species. *Cbln4* is expressed in certain subtypes of turtle SST neurons but not in songbird SST neurons. Data from refs. (Tosches et al., 2018; Colquitt et al., 2021).

The lack of dendrite-dependent bursting in reptiles and songbirds is consistent with comparative electrophysiology within the mammalian brain. Pyramidal neurons in the piriform cortex are homologous to certain types of glutamatergic turtle and songbird neurons (Colquitt et al., 2021) and also lack dendritic plateau potentials (Bathellier et al., 2009). Pyramidal neurons with mammalian electrophysiological properties therefore evolved after interneurons differentiated into PV and SST cell types. (Fig. 5.8).

Ancestral Elfn1 reconstruction

The expression of *Elfn1* by zebra finch and turtle SST-like neurons suggest these cells—and therefore the ancestral SST-like cells— receive(d) facilitating inputs. But it is also possible

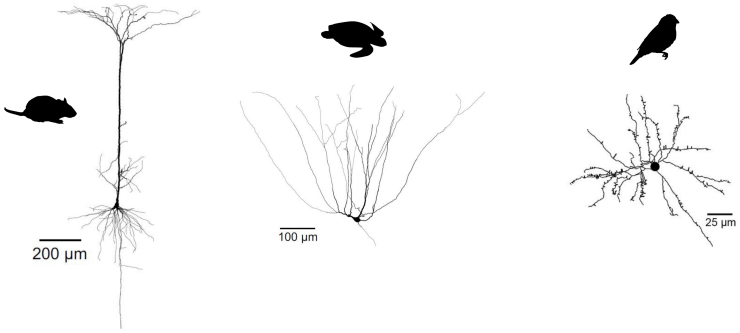

Figure 5.7: Evolutionary divergence of projection neuron morphology. Both turtle and mammalian projection neurons have a pyramidal morphology, but only mammalian pyramidal neurons have a single apical dendrite. Songbird projection neurons have a stellate, not pyramidal morphology. Turtle and mammalian neurons adapted from (Larkum et al., 2008) (published under a Creative Commons License (n.d.)). Songbird neuron adapted from (Kornfeld et al., 2017) (published under a Creative Commons License (n.d.))

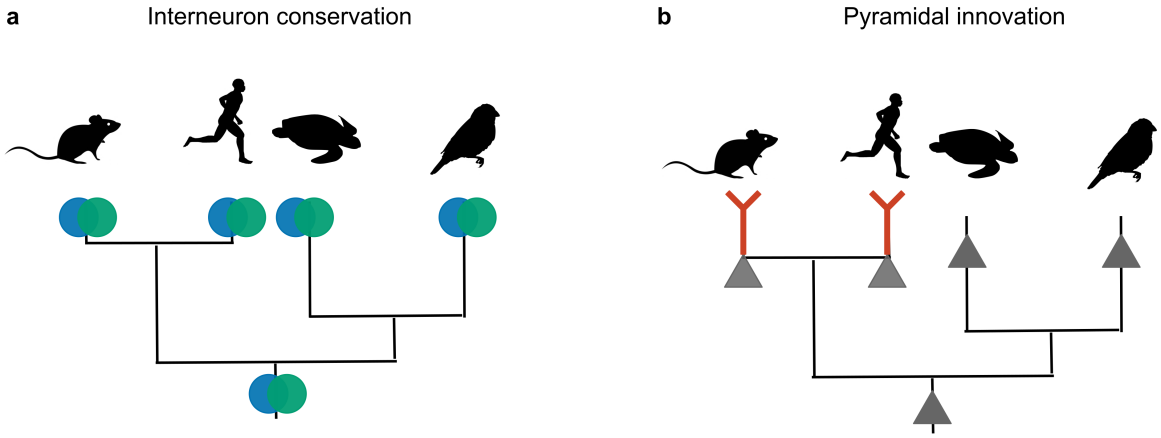

Figure 5.8: Phylogenetic inference of interneuron and pyramidal evolution. (a) mice, humans, songbirds and turtles all have PV and SST interneurons. The most likely explanation for these similarities is that the interneuron types were already present in the last common ancestor of these lineages. **(b)** Only mammalian glutamatergic neurons are known to exhibit dendritic plateau potentials that can elicit burst firing. Other lineages probably lack this trait. The most likely explanation is that dendritic bursting evolved only once in the mammalian lineage.

that the ancestral Elfn1 protein had different functional properties. Previous work has used Elfn1 knockout (Sylwestrak and Ghosh, 2012; Dolan and Mitchell, 2013; Tomioka et al., 2014) and targeted deletions (Dunn et al., 2018) to discover functionally important domains of the mouse variant (Fig 5.9b). To determine the evolution of the Elfn1 protein at the resolution of individual sites, we computationally reconstructed its ancestral state (Hochberg and Thornton (2017) and Orlandi et al. (2023), Methods), starting from the protein sequences of extant species (Fig. 5.9a). Alignment of the extant sequences revealed that, on average, across species, 74.6% of the Elfn1 sites were identical to that of the mouse protein (Fig. 5.9c). Combining the sequence alignment with a probabilistic model of sequence evolution (Jones et al., 1992) and a phylogenetic species-tree allowed us to reconstruct the ancestral protein (Fig. 5.9d). The amount of conservation varied between protein domains and extant species: the zebra finch and turtle sequences were more similar to the ancestral sequence than the mammalian

sequences (Fig. 5.9e). Future work could use the reconstructed sequences to determine the evolutionary history and the molecular mechanisms of short-term facilitation.

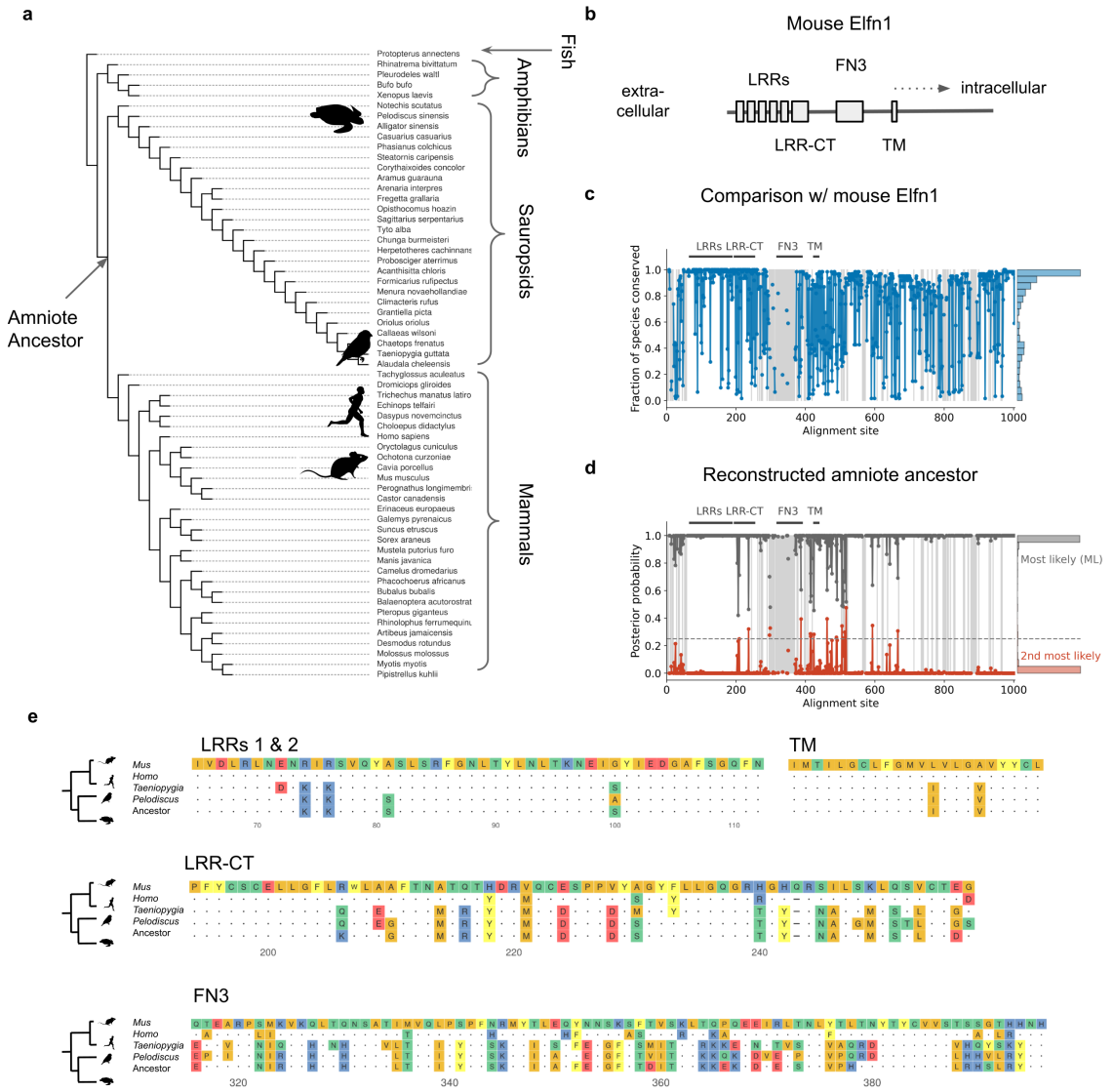

Figure 5.9: Reconstruction of ancestral Elfn1 protein sequences. **(a)** Species-tree showing the phylogenetic relationships of the species whose Elfn1 homologues were used to reconstruct the Elfn1 protein of the amniote ancestor. **(b)** Domain structure of mouse Elfn1 (Dolan et al., 2007; Dunn et al., 2018). LRR: leucine-rich repeat; CT: C-terminal domain; FN3: fibronectin type 3 domain; TM: transmembrane domain. **(c)** Per-site conservation across the tree shown in (a) is computed as the fraction of extant species that share the mouse amino acid in a given site. Dashed lines correspond to gaps. Mean conservation: 0.746. **(d)** Posterior probability of ancestral protein. Grey: most likely (ML) sequence, red: 2nd most likely. Dashed line: cutoff for using the 2nd most likely base in "altAll" sequence. Mean posterior: 0.986. **(e)** Multiple sequence alignment of protein domains shown in (a). Only the first 2 LRRs are shown for space reasons. Dots indicate identity to mouse site; dashes indicate gaps.

5.6 Discussion

The suspicious match between the synaptic properties of PV and SST interneurons and the postsynaptic pyramidal cell compartments suggests that these interneuron properties could

result from an adaptation to pyramidal cells. Here, we evaluated this idea of interneurons being 'fit to function' from an evolutionary and developmental perspective and showed that the relevant interneuron properties predate those of pyramidal cells both during development and in evolutionary history.

Two lines of evidence indicate that the development of PV and SST interneurons is not induced by mature pyramidal cell activity. First, interneurons become committed to a particular cell type (e.g., PV or SST) before reaching the developing cortex. Interneuron fate, therefore, cannot be influenced by the activity of pyramidal cells. Second, at least some of the properties of PV and SST interneurons that strongly shape their control of pyramidal cells — short-term plasticity and output connectivity — emerge before the maturation of pyramidal cell morphology and dendritic activity (dendrite-dependent bursting). It should be noted that other interneuron properties clearly are influenced by pyramidal cell activity. Excitatory activity regulates both the survival of interneurons (Denaxa et al., 2018) and the formation of inhibitory synapses (García et al., 2015; Marques-Smith et al., 2016). Specific types of excitatory neurons determine the laminar allocation of interneurons (Lodato et al., 2011; Wester et al., 2019), and their activity can even change the intrinsic properties of mature interneurons (Dehorter et al., 2015). Cell-extrinsic cues therefore play a role in the normal development of interneurons but are unlikely to determine their identity and the properties we focused on here.

Analogous arguments suggest that the evolution of PV and SST interneurons also cannot be driven by the dendritic physiology of pyramidal cells. The lineages of birds, reptiles and mammals diverged over 300 million years ago, yet they all contain roughly similar interneuron types — evidence that these types were already present in a common ancestor of the three lineages. In contrast to interneurons, excitatory neurons are not conserved and therefore probably evolved later. The second line of evolutionary evidence relates to two specific aspects of interneuron diversity: short-term plasticity and output connectivity. Recent scRNA-seq data (Tosches et al., 2018; Colquitt et al., 2021) show that reptilian and songbird SST interneurons express *Elfn1*, the gene that in mouse SST neurons is necessary and sufficient for short-term facilitation. Certain reptilian, but not songbird, SST subtypes also express the *Cbln4* gene that plays a role in the synaptic specificity of mammalian SST cells (Favuzzi et al., 2019).

These data suggest that ancestral interneurons already comprised PV- and SST-like cell types characterised by some of the genes for cell type-specific phenotypes in mammalian interneurons. It does not, however, imply that these phenotypes were actually present in ancestral cells. The expression of *Elfn1*, for example, is not sufficient for facilitating inputs, as shown in the case of VIP subtypes: Multipolar and bipolar VIP neurons both express *Elfn1*, but only the multipolar subtype receives facilitating excitation (Stachniak et al., 2019). It will therefore be interesting to directly test the presence of PV- and SST-specific phenotypes in reptiles and birds. If neither the reptile nor the songbird homologue of SST interneurons receives facilitating excitatory inputs, *Elfn1* was likely reused for short-term facilitation in mammals. The emergence of short-term facilitation in SST neurons would then be an

adaptation to pyramidal bursting, co-opting pre-existing interneuron diversity for “pyramidal cell purposes". The anatomical connectivity of interneurons might similarly have been reused to control pyramidal cells. In the mammalian brain, PV and SST interneurons inhibit not just the somata and dendrites of pyramidal cells but also of non-pyramidal cells. Ancestral PV and SST interneurons might therefore have specialised in compartment-specific inhibition but not of pyramidal cells for which their presynaptic dynamics are so well-matched.

Although our results show that pyramidal cell bursting is unlikely the driver of the differentiation of PV and SST interneurons, this does not conflict with the functional interpretation of these cell types. In fact, an evolution of active pyramidal cell dendrites before the presence of specialised interneurons would have resulted in aberrant excitation, as seen, e.g., in *Elfn1* mutants (Dolan and Mitchell, 2013; Tomioka et al., 2014). This suggests an alternative picture in which excitatory neurons can only evolve in a way that still allows the existing interneurons to regulate their activity. This still leaves open the question of why interneuron diversity evolved in the first place if it was not for compartment-specific inhibition. Although it is possible that the initial separation between PV and SST cell types was selectively neutral, this is unlikely, given their evolutionary conservation. Instead, the existence of PV and SST cells presumably offers advantages to mammalian and non-mammalian brains alike. An important example of a conserved function could be the temporal coordination of inputs and outputs of pyramidal cells based on oscillations (Klausberger et al., 2003; Bartos et al., 2002).

Our findings have potential implications for the neuroscientific interpretation of optimisation-based models of neural networks, which have recently seen a renaissance (Mante et al., 2013; Yamins et al., 2014; Saxe et al., 2019; Richards et al., 2019; Driscoll et al., 2022). Most of these models describe neural data at the relatively abstract level of dynamics and representations (Kriegeskorte and Diedrichsen, 2019; Vyas et al., 2020). Recently, such deep network models have also started to include circuit-level structures such as separate excitatory and inhibitory populations (Song et al., 2016; Naumann et al., 2022), different neuronal timescales (Kim et al., 2019; Perez-Nieves et al., 2021), and short-term plasticity (Masse et al., 2019; Keijser and Sprekeler, 2022). Deep learning is therefore gradually making its way down from the level of dynamical systems to that of circuits, potentially revealing functional roles for circuit elements. Our findings highlight a challenge to achieving this goal: Multiple circuit-level features — such as the properties of interneuron and pyramidal cells — are interdependent. The function of one feature might depend on that of another and vice versa, raising the question of which features should be optimized (e.g., interneurons) and which should be assumed as pre-existing constraints or opportunities (e.g., nonlinear PC dendrites). In other words, optimization-based models face the challenge of modelling processes such as co-evolution. Merging the functional and evolutionary/developmental perspectives will therefore be an important challenge for future work.

Acknowledgements

We thank Simon J.B. Butt, Loreen Hertäg and members of the Sprekeler Lab for comments on the manuscript.

5.7 Methods

Code was written in Python (version (v) 3.10.8 (vanRossum, 1995)) and R (v4.2.1 (Team, 2021)), based on practices outlined in the Good Research Codebook (Mineault and Nozawa, 2021). Code for the transcriptomic analyses can be found at `https://github.com/JoramKeijser/interneuron_evolution` (JoramKeijser, 2023a). Code for the protein reconstruction can be found at `https://github.com/JoramKeijser/elfn1_reconstruction` (JoramKeijser, 2023b)

Datasets

We analysed the following publicly available single-cell RNA sequencing data sets: Mouse data from Tasic et al. (2018), human data from Bakken et al. (2021), zebra finch data from Colquitt et al. (2021), and turtle data from Tosches et al. (2018). The paper's code repository contains a script for automatically downloading the corresponding files.

For each data set, the starting point of our analysis was a matrix of gene counts per cell, together with the clustering of cells from the original publications. We converted each of the datasets to Seurat (version 4; Hao et al. (2021)) and AnnData (v0.8; Virshup et al. (2021)) objects for downstream analysis in Python and R, respectively. For visual comparison, we labelled songbird and turtle cell clusters according to the most similar mammalian interneuron subclass, as determined in the original publications. This involved the merging of fine-level clusters that presumably capture within-subclass differences. For each dataset, we only visualized cells that were part of, or corresponding to, cortical interneurons. In particular, we did not visualize the correlation of the songbird GABAergic clusters 7, 8, and Pre since these seem homologous to mouse olfactory bulb interneurons (Colquitt et al., 2021).

Dimensionality reduction and clustering

We used AnnData and Scanpy (v1.9.1 (Wolf et al., 2018)) to visualize the expression of the *Elfn1* and *Cbln4* genes. This was done separately for each dataset. We first scaled the counts from each cell to counts per 10 thousand (CP10K) to account for differences in sequencing depth. We then used log plus one pseudo count (log1p) as a variance-stabilizing transformation. Finally, we reduced the dimensionality of each dataset by first finding highly variable genes, performed PCA followed by UMAP (McInnes et al., 2018). We used Scanpy's default parameters for each of these steps. To investigate *Cbln4* expression within the SST population, we performed dimensionality reduction on all SST cells except long-range projecting Chodl cells. Clustering was done using the Leiden algorithm (Traag et al., 2019) with resolution 1.

Correlation analysis

We quantified the overall similarity of species-specific cell clusters by replicating the correlation analysis from Tosches et al. (2018) and Colquitt et al. (2021). We performed the following analysis separately on GABAergic and glutamatergic cells and only compared zebra finch and mice. Specifically, we performed the following steps.

1. Select genes to compare across species. For each species, determine subclass-specific marker genes using Seurat's findAllMarkers (t-test, min.pct = 0.2, max.cells.per.ident = 200) and retain genes with Bonferroni adjusted p-value below 0.05.

2. Intersect the two species-specific lists to find genes that are differentially expressed in both species. This resulted in approximately 500 genes, depending on the cell type.

3. Average counts within each cluster and transform to log scale for variance-stabilization. Specifically, compute: $\log(1 + x) + 0.1$, with x the average count.

4. Divide each gene's value by its average across clusters to obtain a "specificity score" invariant to a gene's overall expression (Tosches et al., 2018).

5. Compute the Pearson correlation between all pairs of mouse and songbird clusters.

We visualised the result using the R package pheatmap (v1.10.12 (Kolde et al., 2012)).

Dataset integration

We used Seurat's anchor-based integration (Stuart et al., 2019) to integrate the zebra finch and mouse data. We did this for GABAergic and glutamatergic neurons separately. First, we jointly performed normalization and variance stabilization for each dataset using Seurat's scTransform (Hafemeister and Satija, 2019), with the percentage of mitochondrial counts as a covariate. Next, we found the top 3000 most variable features across datasets and used these to identify a set of anchors. These were then used to integrate the datasets. Finally, we jointly analysed the integrated datasets using Seurat's standard visualization pipeline: scaling and centring, PCA, and UMAP.

Ancestral Elfn1 reconstruction

We used the Topiary pipeline (Orlandi et al., 2023) to reconstruct the amino acid sequences of the ancestral Elfn1 protein based on sequences of extant species. To this end, we first constructed a source dataset consisting of the Elfn1 sequences from *Mus musculus* (mouse), *Homo sapiens* (human), *Taeniopygia guttata* (zebra finch), and *Pelodiscus sinensis* (Chinese softshell turtle). Next, we used Topiary's seed-to-alignment to find sequence homologs, perform reciprocal BLAST (Altschul et al., 1990) to predict their orthology, reduce sequence redundancy, and align the remaining sequences using Muscle5 (Edgar, 2022). This resulted in 62 aligned sequences used as input to Topiary's alignment-to-ancestors. This infers the maximum likelihood (ML) gene tree, the ML substitution model, and the ML ancestral sequences using RAxML-NG (Kozlov et al., 2019). The posterior probability of an ancestral amino acid was computed using the amino acid's likelihood weighted by its prior probability,

normalised by the sum of all amino acids. Topiary generates bootstrap replicates of the ML gene tree and uses GeneRax (Morel et al., 2020) to reconcile the gene tree with the species tree. The software automatically determined the number of bootstrap replicates to 700. This number—but not the reconstructed ML sequence—varied slightly between runs. Finally, we used topiary's bootstrap-reconcile, which estimates the branch support for the reconciled tree. The ancestral sequence contained 16 ambiguous sites (based on a posterior probability cutoff of 0.25). Besides the ML sequence, we also report a worst-case "altAll" sequence in which the next most likely amino acid has replaced these ambiguous sites. Branch support for the amniote ancestor was 100/100, indicating very high confidence in the existence of this ancestor, as expected. We aligned extant and ancestral sequences using Muscle5 and visualised the resulting alignment using the R package Ggmsa (Zhou et al., 2022).

6

Transcriptomic correlates of state modulation in GABAergic interneurons: A cross-species analysis

6.1 Context within thesis

So far, this thesis has investigated the functions and origins of PV+ and SST+ cells, but interneurons comprise many additional cell types. Recent work found the activity of these fine-grained interneuron types to systematically vary with an animal's behavioural state (Bugeon et al., 2022), at least in mice. Here we test if this phenomenon might generalise across species, as suggested by the homology of interneurons (Ch. 5).

The combination of in vivo imaging and ex vivo transcriptomics from Bugeon et al. (2022) is unavailable in species other than the mouse. We therefore search existing single-cell datasets for transcriptomic correlates of state modulation. We found similar correlates in mice and humans but different correlates across longer evolutionary distances. This indicates that, although homologous, interneurons have acquired lineage-specific expression profiles. Modelling indicates specific ways in which these transcriptomic differences could lead to species-specific state modulation.

6.2 Abstract

GABAergic Inhibitory interneurons comprise many subtypes that differ in their molecular, anatomical and functional properties. In mouse visual cortex, they also differ in their modulation with an animal's behavioural state, and this state modulation can be predicted from the first principal component (PC) of the gene expression matrix. Here, we ask whether this link between transcriptome and state-dependent processing generalises across species. We analysed seven single-cell and single-nucleus RNA sequencing datasets from mouse, human, songbird, and turtle forebrains. Despite homology at the level of cell types, we found clear

differences between transcriptomic PCs, with greater dissimilarities between evolutionarily distant species. These dissimilarities arise from two factors: divergence in gene expression within homologous cell types and divergence in cell type abundance. We also compare the expression of cholinergic receptors, which are thought to causally link transcriptome and state modulation. Several cholinergic receptors predictive of state modulation in mouse interneurons are differentially expressed between species. Circuit modelling and mathematical analyses delineate the conditions under which these expression differences could translate into functional differences.

6.3 Introduction

GABAergic inhibitory interneurons are a highly diverse population consisting of multiple cell types (Bugeon et al., 2022; Markram et al., 2004; Tremblay et al., 2016). In recent years, single-cell RNA sequencing (scRNA-seq; Tang et al. (2009)) has revealed that these types can be further subdivided into tens of subtypes (Zeisel et al., 2015; Paul et al., 2017; Tasic et al., 2018) that also differ in their morphological and electrophysiological properties (Gouwens et al., 2020; Scala et al., 2021). So far, it has been difficult to understand the functional relevance of this fine-grained diversity. Bugeon et al. (2022) recently bridged this gap by revealing that interneurons show subtype-specific modulation with an animal's behavioural state, at least in layers 1-3 of mouse primary visual cortex (VISp). Strikingly, this state modulation could be predicted from the first transcriptomic principal component (tPC1). An interneuron's tPC1 score also correlated with other dimensions of interneuron diversity, such as electrophysiology and connectivity, hinting at an "approximate but general principle" of mouse cortical interneurons (Bugeon et al., 2022).

Intrigued by these findings, we wondered how general the principle embodied by tPC1 actually is (Fig. 6.1). Are transcriptomic correlates of state modulation similar across different species, or at least across mouse cortical layers and areas? If yes, this similarity would suggest conserved principles; if no, the difference could reveal distinct solutions to shared computational problems (Yartsev, 2017; Laurent, 2020; Jourjine and Hoekstra, 2021). The uniformity of interneurons in the mouse brain (Tasic et al., 2018) suggests that their gene expression and state modulation patterns observed in VISp might apply generally. In fact, earlier work by the authors of Bugeon et al. (2022) found that hippocampal interneurons are also organized along a single latent factor (Harris et al., 2018). Similarly, recent comparative transcriptomic analyses have emphasized the conservation of (cortical) inhibitory interneurons across mammals (Hodge et al., 2019; Bakken et al., 2021; Wei et al., 2022; Yu et al., 2023), and more distantly related species (Tosches et al., 2018; Colquitt et al., 2021; Kebschull et al., 2020). But these and other studies (Boldog et al., 2018; Krienen et al., 2020) have also discovered species-specific interneuron subtypes. Additionally, the relative proportions of interneuron types vary even across mouse cortex (Kim et al., 2017; Scala et al., 2019), as does the modulation of interneurons with brain state (Gulledge et al., 2007; Galvin et al., 2018; Khoury et al., 2022; Bratsch-Prince et al., 2023).

We therefore investigated the transcriptomic correlates of state modulation in seven existing single-cell RNA-seq (scRNA-seq) and single-nucleus RNA-seq (snRNA-seq) datasets from the

forebrains of mice (Bugeon et al., 2022; Tasic et al., 2018; Yao et al., 2021), humans (Hodge et al., 2019; Bakken et al., 2021), turtles (Tosches et al., 2018), and songbirds (Colquitt et al., 2021). These species each have homologous types of inhibitory interneurons, but their evolutionary history and brain organization vary in important ways. For example, humans are evolutionarily closest to mice, turtles also have a (three-layered) cortex with potentially different evolutionary origins (Reiner, 1993; Naumann et al., 2015; Tosches et al., 2018), and songbirds lack a cortex altogether (Reiner et al., 2004; Jarvis et al., 2005; Güntürkün and Bugnyar, 2016).

We found that transcriptomic PCs show relatively minor differences across smaller evolutionary distances (e.g., between mice and humans) but diverge over longer evolutionary time scales (e.g., mice and turtles). Between-species differences dwarf within-species differences, likely due to biological rather than technical reasons. Specifically, we trace differences in tPCs to species-specific cell type abundances and within-type gene expression patterns. We also find a combination of conservation and divergence in the expression of the cholinergic receptors correlated with state modulation in mice (Bugeon et al., 2022). Circuit modelling predicts the connectivity patterns for which differences in receptor expression translate into species-specific state modulation of interneurons and cortical information flow.

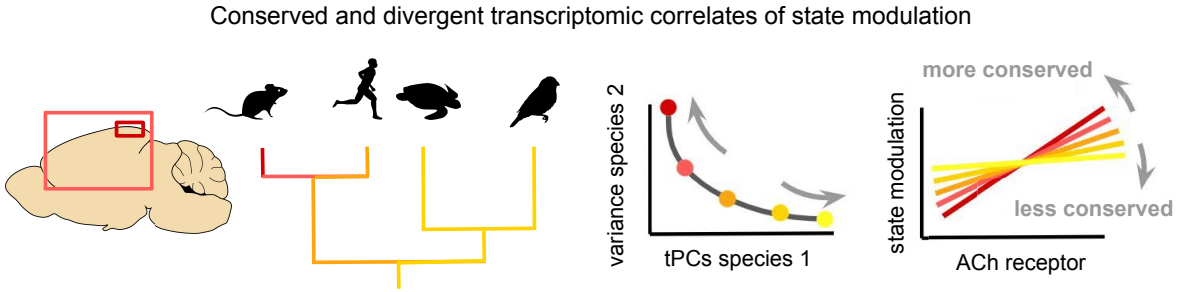

Figure 6.1: Schematic of our main question: could the same transcriptomic axis predict state modulation in other layers and areas of the mouse cortex and in other species? We investigate this by comparing transcriptomic principal components (tPCs) and cholinergic (ACh) receptor expression across published scRNAseq datasets (Bugeon et al., 2022; Bakken et al., 2021; Colquitt et al., 2021; Hodge et al., 2019; Tasic et al., 2018; Tosches et al., 2018; Yao et al., 2021).

6.4 Results

We first validated our analysis pipeline by replicating the relevant results from Bugeon et al. (Bugeon et al., 2022) on their data and conducting several additional analyses. Briefly, we reproduced the systematic variation of interneuron subtypes with behavioural state (roughly, running vs stationary) and its correlation with tPC1 (Fig. B.1). This correlation seems driven by differences within and across cell types (Bomkamp et al., 2019) and is strongest within the Pvalb and Sst populations (Fig. B.2). Whether interneurons form a continuum or cluster along tPC1 depends on the preprocessing of the transcriptomic data (Fig. B.3). These caveats aside, our analyses are consistent with those from Bugeon et al. and might reveal similar patterns—or the lack thereof— in other brain areas and species. A detailed description of the replication can be found in the supplementary material (see Replication of Bugeon et al).

Conserved and divergent transcriptomic axes across species

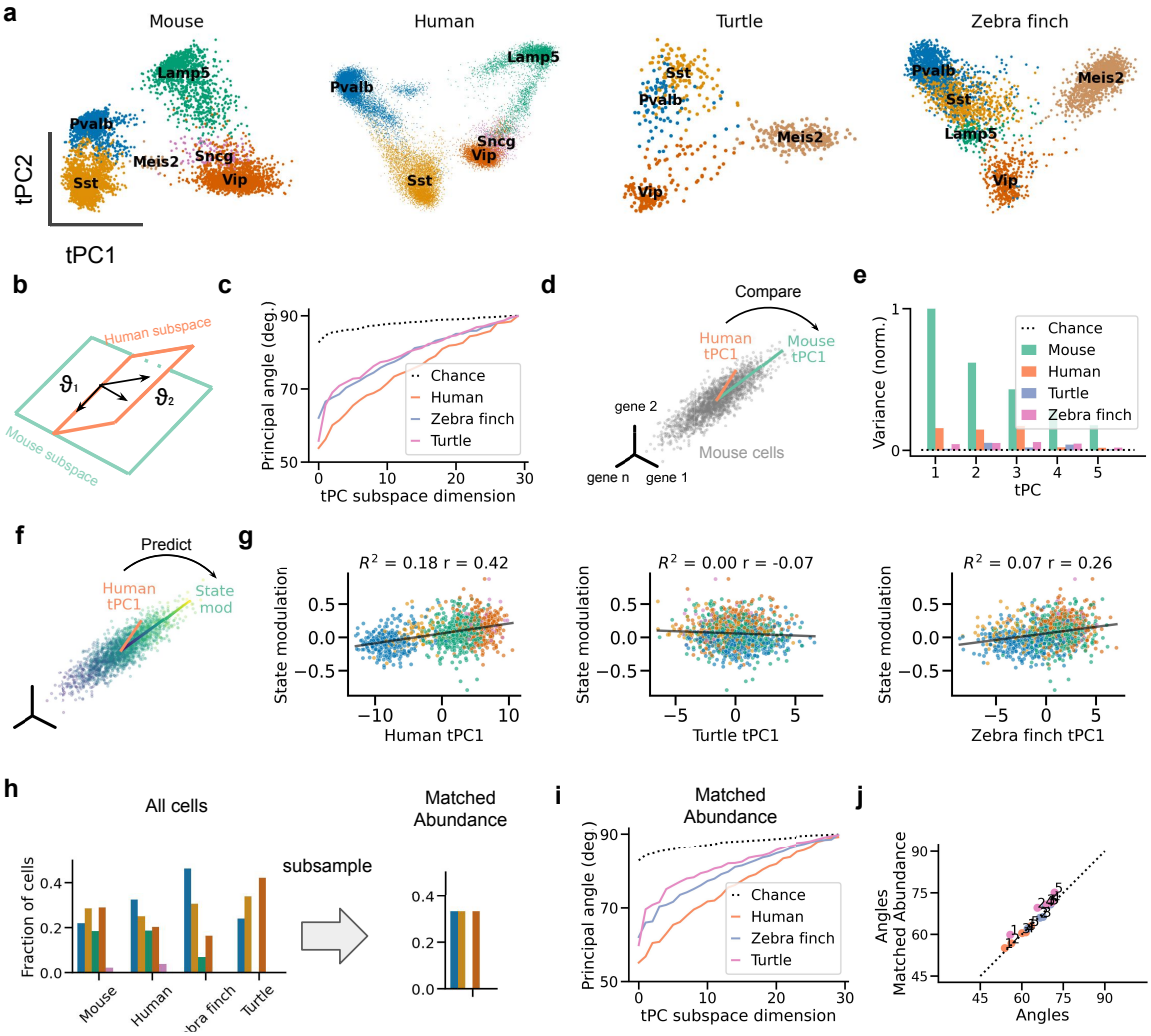

Figure 6.2: Transcriptomic PCs capture conserved and divergent global gene expression patterns. **(a)** Projections of gene expression data from forebrain interneurons onto each dataset's first tPCs. Samples correspond to cells; colours indicate cell types. Variance explained by tPC1&2: 23.3% (mouse), 22.7% (human), 13.7% (turtle), and 12.9% (zebra finch). **(b)** Schematic: principal angles measure dissimilarity between subspaces spanned by two sets of tPCs (see Methods). Larger angles indicate larger differences. **(c)** Principal angles between human and mouse subspaces are smaller than between mouse and zebra finch or turtle subspaces. Chance level estimated by sampling random, normalized vectors. **(d)** Schematic: Variance explained in the mouse data as a measure of tPC similarity. A tPC's length is proportional to the variance it explains. **(e)** Variance of mouse data explained by tPCs of different datasets, normalized to the variance explained by mouse tPC1. The human, zebra finch, and turtle tPC1 explain 15.7%, 4.5%, and 1.2% of the variance explained by mouse tPC1. A random direction (dashed line) explains 0.3%. **(f)** Schematic: Predicting mouse state modulation from human tPC1. The colour gradient symbolizes the state modulation of mouse cells. **(g)** State modulation of mouse interneurons can be predicted from the interneuron's projection onto human tPC1 but not onto turtle or zebra finch tPC1. R^2: cross-validated fraction of variance explained, r: Pearson correlation. **(h)** Subsampling procedure to control for the relative abundance of interneuron subclasses across datasets. Colours code for cell types (see (a)). **(i,j)** Matching the relative abundance does not increase the similarity of datasets as measured using principal angles. Numbers indicate order or PCs. Data from Tasic et al. (2018) (mouse), Bakken et al. (2021) (human), Tosches et al. (2018) (turtle), Colquitt et al. (2021) (zebra finch).

Having validated our approach on mouse data, we next turn to the cross-species comparison using transcriptomic data from humans (*Homo sapiens*, Bakken et al. (2021)), turtles (*Trachemys scripta elegans*, Tosches et al. (2018)), and zebra finch (*Taeniopygia guttata*, Colquitt et al. (2021)); see Table 6.1 for an overview of all analysed datasets. We compare these data with a large reference dataset from mouse VISp (Tasic et al., 2018).

We first visualized the data from different species. To this end, we preprocessed the datasets using the same analysis pipeline and applied PCA to the resulting RNA count matrices (see Methods). The projection onto the first 2 tPCs of the human, but not turtle or zebra finch data, was similar to that of the mouse data (Fig. 6.2a). Mouse and human interneurons clustered by developmental area (Lim et al., 2018a), with medial ganglionic eminence (MGE)-born Pvalb and Sst cells occupying one side of tPC1, and caudal ganglionic eminence (CGE)-born Lamp5, Vip, and Sncg cells the other. An intermediate position was occupied by a small group of Meis2 neurons (Tasic et al., 2018), located in the white matter (Frazer et al., 2017). In contrast to the mammalian datasets, the turtle and finch data were characterised by a large population of Meis2-positive neurons (Fig. 6.2a, Table 6.2). Transcriptomic and morphological evidence suggests that these cells are likely homologous to neurons in the mammalian striatum rather than the white matter (Tosches et al., 2018; Colquitt et al., 2021).

We quantified these visual differences using the principal angles, which generalise the notion of angle between two lines in a plane (Fig. 6.2b; see Methods). Here, we computed the angles between subspaces spanned by each dataset's top 30 PCs. Consistent with the impression from the first 2 PCs, the principal angles were the smallest between mouse and human subspaces (Fig. 6.2c). Turtle and zebra finch PCs were both dissimilar to mouse PCs. Principal angles do not require a one-to-one relationship between individual principal components but also do not consider the variance explained by these components. For instance, a pair of highly similar but low-variance dimensions will result in small principal angles—inadvertently suggesting high similarity. We therefore performed a complementary analysis by computing the variance in the mouse data explained by the PCs of other datasets (Fig. 6.2d). The first human PC accounted for 15% of the variance explained by the first mouse PC; the turtle and songbird tPC1 accounted for 1.2% and 4.5%, respectively (Fig. 6.2e). Each tPC1 explained more variance than a random direction (0.3%), consistent with some shared global structure.

We confirmed that these results were not due to technical differences in the different datasets. We first controlled for sequencing depth using a subsampling procedure (Fig. B.4, Methods). We also mapped each dataset onto the mouse data using anchor-based integration (Stuart et al., 2019). This method has been widely used in cross-species analyses (e.g., Hodge et al. (2019), Colquitt et al. (2021), Hain et al. (2022), and Keijser and Sprekeler (2023)). As expected, computational integration increased the similarity among the datasets (Fig. B.5), but the larger similarity between human and mouse data was preserved.

How might the transcriptomic differences relate to state modulation? Because state modulation information was only available for the mouse (Bugeon et al., 2022), we projected this data onto the tPCs from other datasets to determine their predictive ability (Fig. 6.2f). We found that the human tPC1 predicts state modulation in the mouse ($R^2 = 0.18$), but the turtle tPC1 did not ($R^2 = 0$) (Fig. 6.2g; compare with Fig. B.1). The zebra finch tPC1 showed a weak but significant ability to predict state modulation ($R^2 = 0.07$). We conclude

that human tPCs are similar to those of the mouse also on a functional level, in line with evolutionary history.

What evolutionary changes underlie the differences between transcriptomic PCs? At least two non-mutually exclusive processes are possible. First, homologous subclasses could evolve in a species-dependent manner, as indicated by differences in gene expression. Second, evolution can also change the relative abundance of otherwise conserved classes (Bakken et al., 2021; Chartrand et al., 2023). We wondered if the relative abundance of cell classes was sufficient to explain the species differences. To this end, we resampled cells to equal fractions, such that the 3 classes (Pvalb, Sst, Vip) present in all datasets each accounted for one-third of the cells (Fig. 6.2h). This increased the visual similarity between the first two tPCs of the mammalian and non-mammalian datasets due to the absence of Meis2 neurons (Fig. B.6). Still, the matched-abundance datasets were as dissimilar as the original datasets (Figs. 6.2j, B.7). This suggests a divergence between homologous cell types as a driver of evolutionary change in the global transcriptomic landscape.

The pronounced differences in gene expression patterns may be surprising given the known homology of GABAergic interneuron subclasses, which contrasts with the evolutionary changes in glutamatergic cells (Tosches et al., 2018; Colquitt et al., 2021; Keijser and Sprekeler, 2022). We therefore wondered if the transcriptomic PCs of glutamatergic cells would be even more dissimilar between species. We found that is indeed the case, but the differences are small: the PCs of glutamatergic neurons are not much more species-specific than those of interneurons (Fig. B.9). These analyses show that, despite a common evolutionary history, homologous GABAergic cells have undergone lineage-specific changes that are reflected not only among fine subtypes but also in global patterns of gene expression.

Similar transcriptomic axes across mouse datasets

The previous cross-species comparison is based on data collected with different sequencing protocols and from different brain areas. To account for these factors, we calibrated the between-species differences against within-species differences by comparing three mouse datasets (Fig. 6.3a): the in situ data from VISp layers (L) 1-3 (Bugeon et al., 2022), the plate-based (SMART-seq2) data from VISp L1-6 (Tasic et al., 2018), and the droplet-based (10X) data from multiple cortical and hippocampal areas (Ctx & Hpc, Yao et al. (2021))

Visually, the projections onto the first tPCs were similar (Fig. 6.3b), with interneurons clustering by developmental area, as before. But subtle differences were also visible. For example, the L1-3 dataset lacked a Meis2 population present in both L1-6 datasets (Fig. B.8; (Tasic et al., 2018; Yao et al., 2021)). The tPC1 score of Sst cells also varied between datasets. In the L1-3 data, Sst cells occupied an intermediate position on tPC1 (and tPC2) compared to Pvalb cells, consistent with their weaker state modulation (Fig. B.1b). In contrast, Sst and Pvalb cells occupied similar positions in the other datasets.

The three datasets were also quantitatively similar. Principal angles between different mouse datasets were substantially smaller than angles between species (Fig. 6.3c). To compare the larger mouse datasets with the smaller dataset of Bugeon et al. (2022), we performed the same analyses after selecting the 72 genes shared by all datasets. This revealed the Ctx & Hpc

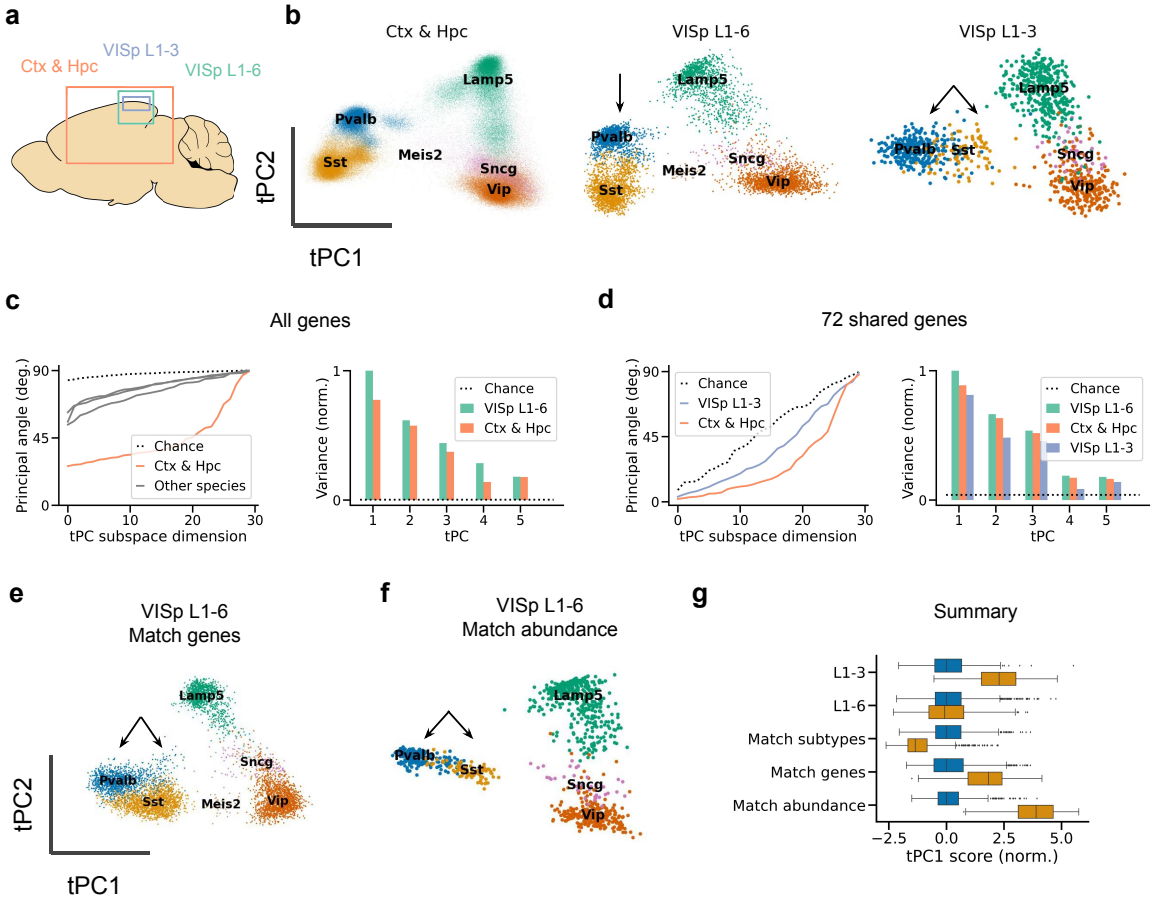

Figure 6.3: Similar transcriptomic PCs across mouse data sets. **(a)** Schematic of brain areas sequenced for different mouse datasets. Ctx: cortex, Hpc: hippocampal formation, VISp: primary visual cortex. The datasets were also collected using different technologies (Table 6.1). **(b)** Similar tPC1&2 across mouse datasets; tPCs1&2 jointly explain 18.0%, 23.8%, and 29.8% of variance, from left to right. Arrows indicate a qualitative difference: the relative position of Pvalb and Sst cells along tPC1. **(c)** Quantitative comparison with VISp L1-6 dataset based on 2000 highly variable genes; tPC1 of the Ctx & Hc dataset explains 77.5% of the variance explained by tPC1 of VISp L1-6. Grey lines: cross-species angles, taken from Fig. 6.2c. **(d)** As (c), but based on the 72 genes shared by the three datasets. In this reduced space, tPC1 of the Ctx & Hc and VISp L1-3 explains 88.7% and 81.3%, respectively, of the variance explained by VISp L1-6. **(e)** Relatively small change in Pvalb and Sst position after matching gene sets between L1-6 and L1-3 data. **(f)** Larger differences due to relative cell type abundance. **(g)** Distribution of tPC1 projection of Pvalb (blue) and Sst (orange) cells for the L1-3 data and different versions of the L1-6 data. Match subtypes: select only the Sst subtypes present in the L1-3 dataset. Match genes: select only the genes present in the L1-3 dataset. Match abundance: subsample such that Sst cells comprise only 8% of the samples, as in the L1-3 data. Projections were normalized such that the mean and variance of the respective Pvalb population were zero and one, respectively. Expression data from Yao et al. (2021) (Ctx & Hpc), Tasic et al. (2018) (VISp L1-6), and Bugeon et al. (2022) (VISp L1-3).

data to be more similar to the VISp L1-6 data than the VISp L1-3 data (Fig. 6.3d), consistent with the varying relative positions of the cell types in the space of the first two PCs.

Several factors could explain the different positions of Pvalb and Sst cells along tPC1 and tPC2 (arrows in Fig. 6.3b). We first tested if the difference was due to layer-specific subtypes known to be transcriptomically identifiable (see, e.g., Tasic et al. (2018) and Wu et al. (2023)). However, selecting L1-3 subtypes from the L1-6 data followed by PCA only moved the Sst cells further along tPC1 (Fig. 6.3g, "match subtypes"). We next tested for the influence of

gene set by performing PCA on the L1-6 data after selecting 72 genes describing the L1-3 dataset. This only modestly increased the similarity to the in situ data (Fig. 6.3e), reflecting the careful selection of the gene panel (Qian et al., 2020). Finally, we reasoned that the intermediate position of Sst cells in the L1-3 data could be due to their relative sparsity (8% in the L1-3 data vs 28% in the L1-6 data). After all, a given pattern of covariability explains less variance when present in a smaller number of samples. Indeed, sampling the same number of cells from the entire Sst population moved the Sst population to an intermediate tPC1 position (Fig. 6.3f). Therefore, the intermediate position of Sst cells in the L1-3 dataset might be due to their relative sparsity.

In summary, mouse datasets are highly similar to cross-species datasets despite differences in brain area (Tasic et al., 2018) and sequencing technology (Bakken et al., 2018). Two human datasets (Hodge et al., 2019; Bakken et al., 2021) showed equally high levels of similarity (Fig. B.10). Between-species differences, therefore, likely reflect biologically meaningful signals rather than technical artefacts.

Evolution of cholinergic receptor expression

So far, we have shown that interspecies expression differences are reflected in the first principal components. This rules out a conserved tPC1 that predicts state modulation—at least across evolutionarily distant species. However, it does not rule out that the species-specific tPCs predict state modulation. Unfortunately, this cannot be tested directly due to the lack of data on state modulation for the other species. As a proxy, we, therefore, analyzed the expression of cholinergic receptors that are known to contribute to the correlation between tPC1 and state modulation in mice (Fig. 6.4a; Bugeon et al. (2022)).

According to our analysis (see Methods), five cholinergic receptors can predict state modulation of upper-layer subtypes in held-out data in mice (Fig. B.11). These receptors also ranked among the top genes in their correlation with state modulation (Fig. 6.4b). The predictive nicotinic receptors (Chrna3,4,5) showed a rough gradient along tPC1 (see, Fig. 6.4c). The only predictive inhibitory receptor (Chrm4), on the other hand, was expressed by Pvalb neurons (Figs. 6.4d), consistent with their negative state modulation (Fig. B.1b).

Do the same receptors mediate state modulation in other species? If yes, one would expect differential expression across cell types, with a similar pattern as in mice. However, several receptors that predict state modulation in mice show qualitatively different patterns of expression in the other species (Fig. 6.4e,f,g). For example, Chrna4 and Chrna5 show much weaker expression in the human data than in mice (Fig. 6.4e). Chrm4 is overexpressed in the turtle data relative to the other species (Fig. 6.4f).

The general trend is that the predictive receptors are under-expressed in the other datasets. A possible explanation is a regression to the mean: predictive receptors are, by necessity, expressed in mice. But the relative expression in other datasets could also be due to technical reasons such as a lower sequencing depth (Table 6.1). Indeed, the typical mouse cell contained several orders of magnitude more RNA counts than the typical human cell (Fig. 6.4h). We controlled for this confound by downsampling the mouse data to the sequencing depths of the other datasets (Methods). To measure variability, we also applied this procedure to

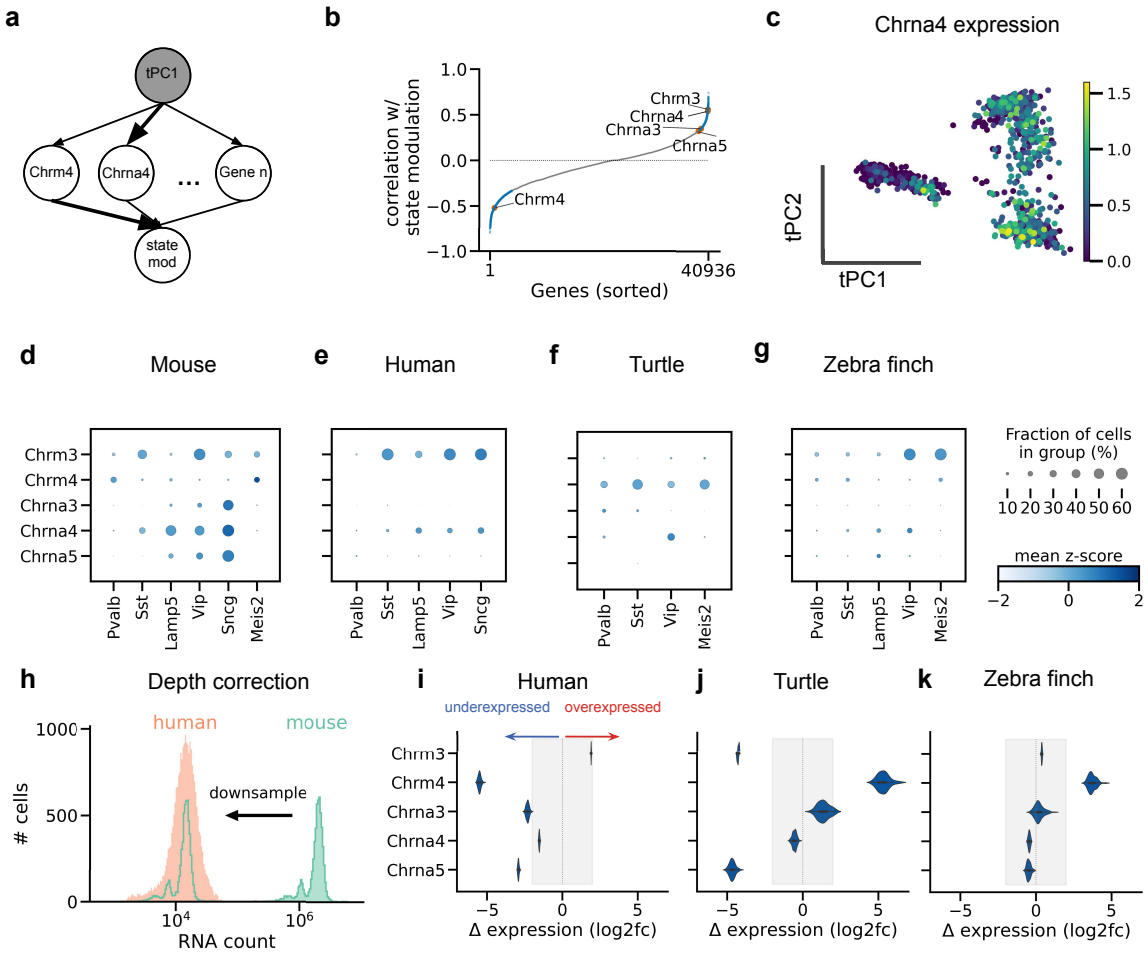

Figure 6.4: Evolution of cholinergic receptor expression. **(a)** Graphical model of the relationship between tPC1, gene expression, and state modulation. The shaded variable tPC1 is not directly observable. Arrows indicate direct dependence relationships; arrow thickness indicates the strength of the relationship. **(b)** Correlation between gene expression and state modulation in the mouse data. Gene expression and state modulation were not measured in the same cells and were therefore combined at the level of subtypes (Methods). Blue: significant correlation ($p < 0.05$), grey: not significant. Annotated are the cholinergic receptors that predict state modulation (Fig. B.11). Chrm4 is ranked 327th (top 1.5%) with the strongest negative correlation. Chrm3 and Chrna3, 4 & 5 are among the top 1.7%, 2.4%, 7.1%, and 8.1% with the strongest positive correlation. **(c)** tPC projection of mouse VISp L1-6, coloured by Chrna4 expression (log CP10K). Upper layer Sst types were selected before PCA to retain the cell type arrangement of Bugeon et al. (Fig. 6.3f) **(d-g)** Dotplots of cell type-specific cholinergic receptor expression, z-scored across all cells. **(h)** Schematic of RNA count subsampling to control for differences in sequencing depth. Each RNA count from the deeper dataset was sampled with a probability equal to the relative depth of the deep and the shallower dataset (Methods). **(i-k)** Log2-fold difference in expression with mouse data after subsampling; negative and positive values indicate under- and overexpression, respectively, compared to mouse data. Each violin plot shows the distribution of 100 subsampled datasets. Differences outside of shaded areas are larger than the typical differences between different datasets of the same species (Fig. B.12). Without the subsampling procedure, a comparison of raw RNA counts would suggest that every receptor is overexpressed in the mouse by a factor of 8 or more due to the larger sequencing depth of the mouse data. Expression data from Tasic et al. (2018) (mouse), Bugeon et al. (2022) (VISp L1-3), and Tasic et al. (2018) (L1-6).

two datasets from the same species, which revealed typical log2-fold expression differences between -2 and 2 (Fig. B.12)—downsampling retained larger differences between species that are qualitatively consistent with the analysis of the full datasets. In the human data, Chrm4,

and to a lesser extent Chrna3 & 4, were still underexpressed after downsampling (Fig. 6.4i). Chrm3 and Chrna5 were underexpressed in the turtle data, whereas Chrm4 was overexpressed (Fig. 6.4j). In the songbird data, only Chrm4 was overexpressed (Fig. 6.4k).

Thus, several cholinergic receptors that might mediate state modulation in mice show species-specific expression. This suggests that homologous cell types in different species could show substantial differences in state modulation.

Robustness of state modulation to cholinergic receptor expression

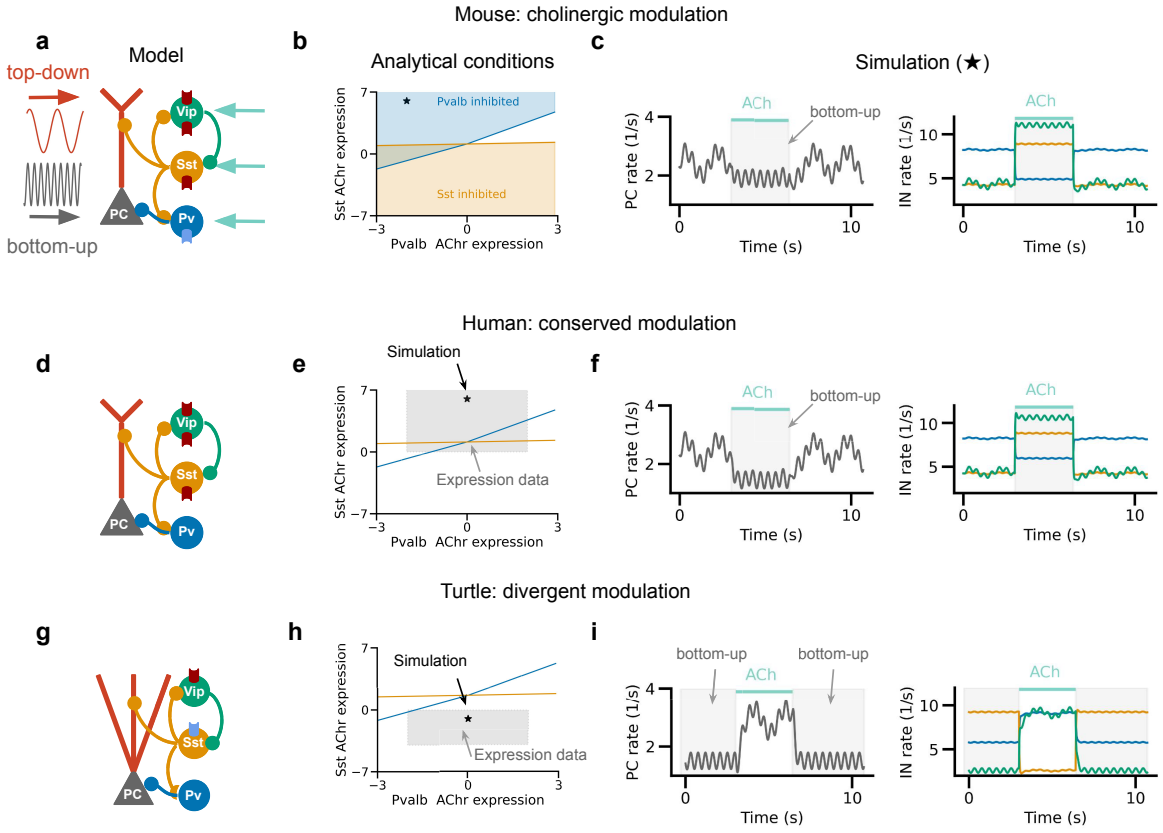

Figure 6.5: Circuit model predicts functional consequences of cholinergic receptor expression divergence (a) Schematic of mouse cortical circuit model. PC: pyramidal cell. Grey and red sinusoids represent inputs to pyramidal soma and dendrites, respectively; teal arrows represent cholinergic modulation of interneurons. Excitatory and inhibitory ACh receptors are shown in red and blue, respectively. **(b)** Analytical prediction of the Pvalb and Sst ACh receptor expression values for which Pvalb and Sst cells are inhibited by ACh (see Methods). The star indicates parameter settings used for simulation in **(c)**. At baseline, the PC rate reflects both somatic and dendritic inputs. Cholinergic activation inhibits the dendritic contribution by activating Sst cells. **(d)** Human cortical circuit model, in which Pvalb cells lack inhibitory ACh receptors. **(e)** As (b), but with the shaded area indicating expression values qualitatively consistent with the human data. Star: parameter settings used for simulation in **(f)**. Cholinergic modulation has a qualitatively similar effect due to the indirect cholinergic inhibition of Pvalb neurons via the Sst neurons. **(g)** Turtle circuit model with inhibitory ACh receptor expression in Sst cells. **(h)** As (b), but with the shaded area indicating turtle expression values. Star: parameter settings used for simulation in **(i)**. In contrast to the mammalian circuit models, cholinergic modulation causes a disinhibition of dendritic inputs.

How do species-specific cholinergic receptor expression patterns influence cortical information flow? Since this depends not just on the cell type-specific gene expression but also on the interplay of different interneurons, we investigated this question using a circuit model (Methods). We focused on the most salient differences in receptor expression between the three species with a cortex (mice, humans, and turtles).

The model consists of the three most common interneuron types, Pvalb, Sst, and Vip cells, whose connectivity patterns have been mapped (Pfeffer et al., 2013; Jiang et al., 2015) and are relatively conserved, at least in mice and humans (Campagnola et al., 2022). Additionally, the computational repertoire of this "canonical circuit" has previously been investigated (Litwin-Kumar et al., 2016; Yang et al., 2016; Hertäg and Sprekeler, 2019). To explore the effect of cholinergic modulation on excitatory activity, we also included a two-compartmental pyramidal neuron. These two compartments receive different information streams: whereas the soma receives feedforward (sensory) input, the pyramidal dendrites receive top-down input (Larkum, 2013; Ju et al., 2020; Schroeder et al., 2023). For visualisation purposes, these input streams were represented by sinusoids of different frequencies (Fig. 6.5a).

We model cholinergic modulation as an additive input to the interneurons with a strength that is based on cell type-specific receptor expression data (Methods, Table 6.3). Cholinergic modulation inhibits the Pvalb population via muscarinic ACh receptors (Chrm4) while activating Vip cells and — to a lesser extent — Sst cells via nicotinic ACh receptors. The activated Sst cells suppress inputs arriving at the dendrites (Murayama et al., 2009; Silberberg and Markram, 2007; Keijser and Sprekeler, 2022), and increase the effective cholinergic inhibition of Pvalb cells. By recruiting dendritic inhibition, ACh therefore limits the influence of top-down inputs on PC rates (Parikh et al. (2007) and Guillem et al. (2011); Fig. 6.5c) while enhancing the influence of top-down inputs (Hasselmo, 1995; Moran et al., 2013).

How might cholinergic modulation affect interneuron activity and information flow in other species? We modelled a human cortical circuit by deleting the inhibitory Chrm4 receptor from the Pvalb population, mimicking the strongest difference with the mouse VISp data (Fig. 6.5d). Mathematical analyses indicate that Pvalb activity is still reduced in the presence of ACh due to inhibition from Sst cells (Fig. 6.5e,f; see Methods). The difference in direct cholinergic inhibition in Pvalb interneurons between humans and mice might therefore have a relatively weak functional consequence. This is consistent with experimental data (Chen et al., 2015b) and the variable Chrm4 expression across mouse datasets (Fig. B.12). By contrast, we found a qualitatively different effect after changing the ACh receptor densities to mimic the turtle data (Fig. 6.5g): the over-expression of inhibitory Chrm4 receptors by Sst instead of Pvalb cells led to cholinergic disinhibition of dendritic inputs (Fig. 6.5h,i). This qualitative deviation from the mammalian cortex could affect the state-dependent processing of bottom-up versus top-down inputs in reptiles (see Discussion).

These findings suggest that the computation performed by the circuit can be very sensitive to certain patterns of differential expression and robust to others (Prinz et al., 2004; Marder et al., 2014).

6.5 Discussion

We have shown that the global gene expression patterns of inhibitory interneurons, as assessed by PCA, show considerable similarity between mice and humans. However, such similarity is not observed between mice and turtles or songbirds. This suggests that the first transcriptomic PC (tPC1) obtained from the upper layers of the mouse cortex (Bugeon et al., 2022) may serve as an organizing principle for mammalian interneurons but not for reptilian and avian interneurons. Control analyses indicate that technical factors cannot explain cross-species differences. Instead, our results suggest that the evolutionary divergence of homologous interneurons is mainly explained by changes in gene expression rather than changes in the relative abundance of cell types. Alongside the differences in global expression patterns, we have also observed differences in the expression of cholinergic receptors, suggesting that interneurons undergo species-specific modulation in their functional states.

Comparison to prior work

The gross transcriptomic differences between species might be surprising given the evolutionary conservation of interneurons in the forebrain (Tosches et al., 2018; Colquitt et al., 2021; Bakken et al., 2021; Wei et al., 2022) and other areas (Kebschull et al., 2020; Yu et al., 2023). However, these works also found many genes to have species-specific expression, suggesting cell types might be homologous across species but not preserved in their detailed properties. Moreover, fine interneuron subtypes are not necessarily conserved across larger evolutionary distances (Tosches and Laurent, 2019). Our analyses reveal global differences in gene expression between homologous cell types that rival the differences between glutamatergic cells (Fig. B.9).

Cholinergic modulation with arousal and other cognitive processes has been reported in many species (see, e.g., Disney et al. (2007), Xiang et al. (1998), Pinto et al. (2013), Zhu et al. (2023), Puzerey et al. (2018), and Jaffe and Brainard (2020)), and might even be mediated by similar midbrain cell types (Lovett-Barron et al., 2017). However, acetylcholine seems to achieve its conserved effects via species-specific pathways (Coppola and Disney, 2018). For example, most human but not rat PV neurons express the Chrm1 receptor (Disney and Reynolds, 2014). Even within a single species, cholinergic projections and their effects vary across areas (Avendano et al., 1996; Li et al., 2018; Khoury et al., 2022) and layers (Poorthuis et al., 2013; Obermayer et al., 2017). Additionally, serotonergic receptors are among the most differentially expressed gene sets between humans and mice (Hodge et al., 2019). Therefore, our differential receptor expression findings are broadly consistent with earlier work. It should be noted, however, that technical differences between datasets naturally pose a more serious limitation for comparing the expression of individual genes than aggregate measures such as PCA. Future work will therefore need to confirm the present results using, e.g., immunohistochemistry.

Interpretation of transcriptomic PCs

Across all datasets, one feature consistently stands out: the clustering along tPC1&2 by developmental area. A similar pattern has been previously observed based on nonlinear dimensionality reduction and clustering methods (see, e.g., Tasic et al. (2018) and Keijser

and Sprekeler (2023)). The structuring of top PCs by developmental origin and cell type is expected since cell types are defined by developmentally-activated transcription factors that coregulate batteries of protein-coding genes (Hobert, 2008; Arendt et al., 2016). These low-dimensional patterns of gene expression are naturally picked up by a method like PCA.

A clear difference between the mouse datasets is given by two layer-specific subtypes: the deep-layer Meis2 cells (Frazer et al., 2017; Tasic et al., 2018), and long-range projecting Sst Chodl cells (Melzer and Monyer, 2020; Tasic et al., 2018). Their intermediate position along tPC1 (Fig. B.8) but distinct connectivity suggests that the correlation of tPC1 with cellular properties (Bugeon et al., 2022) might not apply to these deep layer subtypes. This could be further tested using, for example, Patch-seq experiments (Cadwell et al., 2016; Lipovsek et al., 2021). A caveat is the relative sensitivity of PCA to cell-type proportions: the intermediate tPC1 scores of Meis2 cells may be caused by their scarcity in the mouse data.

Evolution of cholinergic modulation

Many genes are at least as predictive of state modulation as cholinergic receptors (Fig. 6.4b). Some genes might be causally related to state modulation, but others are merely co-regulated with causal genes (Fig. 6.4a). Thus, strong correlations—the very property that allows the reliable identification of transcriptomic PCs—also preclude the identification of causal genes based on regression analyses. Whether genes predict state modulation also depends on factors only partially under genetic control, such as synaptic connectivity. For example, our network simulations show that connectivity patterns influence which cholinergic receptor expression differences affect state modulation (Fig. 6.5). Finally, species differences in state modulation may not necessarily imply differences in function. An interesting example is turtle Sst cells' expression of the inhibitory Chrm4 receptor, which might lead to cholinergic disinhibition of pyramidal dendrites. Since sensory inputs to turtle cortex arrive in layer 1 instead of deeper layers (Reiner, 1993; Ulinski, 2007), we speculate that acetylcholine can thus disinhibit sensory inputs, as it potentially does in mouse cortex. Alternatively, cholinergic modulation could have qualitatively different effects on the processing of bottom-up versus top-down inputs in turtle compared to mouse cortex. Future experiments could arbitrate between these alternatives.

Conclusion

The wide availability of transcriptomic data in different species offers new opportunities for comparative analyses. Transcriptomic data can not only predict behavioural features such as state modulation but also the electrophysiology and morphology of homologous cell types (Tripathy et al., 2017; Gouwens et al., 2020; Bernaerts et al., 2023), which are more accessible. It will be exciting to see whether these predictions generalize across species and if they correlate with high-variance transcriptomic dimensions. More generally, we expect that future cross-species experiments will complement work in genetically accessible mice to reveal general principles of brain function.

Acknowledgement

We thank Trygve Bakken, Michael Brainard, Stéphane Bugeon, Bradley Colquitt, Kenneth Harris, and Bosiljka Tasic for their feedback. We are also grateful to Sadra Sadeh for suggesting a modelling approach.

6.6 Methods

Code was written in Python and R and combined into a reproducible workflow using Snakemake (Köster and Rahmann, 2012). The code will be made available at `https://github.com/JoramKeijser/transcriptomic_axes` upon publication.

Datasets

An overview of the analysed datasets is shown in Table 6.1, Table 6.2 lists the relative frequency of different cell types in each dataset.

Name	Species	Area(s)	Technology	Cells	Genes/cell	Reads
Bugeon (Bugeon et al., 2022)	Mus musculus	VISp L1-3	CoppaFish (sc)	1,065	49	-
Tasic (Tasic et al., 2018)	Mus musculus	VISp	SS v4 (sc)	6,125	9,795	2,009,806
Yao (Yao et al., 2021)	Mus musculus	Ctx & Hpc	10x V2 (sc)	177,614	3,750	8,957
Bakken (Bakken et al., 2021)	Homo sapiens	M1	SS v4 & 10X (sn)	23,992	4,719	14,337.5
Hodge (Hodge et al., 2019)	Homo sapiens	MTG	SS V4 (sn)	4,164	8,344.5	1,901,796.5
Colquitt (Colquitt et al., 2021)	Taeniopygia guttata	HVC & RA	10X v2 (sn)	3,786	161.5	3,150.5
Tosches (Tosches et al., 2018)	Trachemys scripta elegans	Ctx	Drop-Seq v3 (sc)	640	2,952	6,284

Table 6.1: Overview of analysed datasets. Sn: single nucleus, sc: single cell. SS: Smart-Seq. M1: primary motor cortex, VISp: primary visual cortex, HVC: high vocal centre, MTG: middle temporal gyrus, Ctx: cortex, Hpc: hippocampal formation. Genes/cell: median number of genes detected per cell. Reads: median number of reads per cell. For the single nucleus data, reads aligned to exons and introns were used. The data from Bugeon et al. (2022) comprised already normalized counts; the number of reads was therefore unavailable.

	Bakken (Bakken et al., 2021)	Bugeon (Bugeon et al., 2022)	Colquitt (Colquitt et al., 2021)	Hodge (Hodge et al., 2019)	Tasic (Tasic et al., 2018)	Tosches (Tosches et al., 2018)	Yao (Yao et al., 2021)
Pvalb	32.4	27.9	31.9	17.5	21.8	13.8	17.2
Sst	25.0	8.5	21.0	30.6	28.4	19.5	26.7
Lamp5	18.6	35.8	4.7	27.8	18.3	0.0	23.7
Vip	20.2	24.4	11.2	24.1	28.7	24.2	24.6
Sncg	3.7	3.5	0.0	0.0	2.0	0.0	7.8
Meis2	0.0	0.0	31.3	0.0	0.7	42.5	0.0

Table 6.2: Percentage of cell types for each dataset, rounded to a single decimal place. The mammalian Meis2-positive cells are likely not homologous to the turtle/finch Meis2-positive cells ((Colquitt et al., 2021), see text) but are grouped for convenience.

Replication of Bugeon et al.

The starting point of our replication was the in vivo calcium imaging data and in situ transcriptomic data previously described by Bugeon et al. (2022). We preprocessed and analysed these data following the original paper unless indicated otherwise. We selected interneurons with a high-confidence assignment to a particular subtype (posterior probability at least 0.5; Qian et al. (2020) and Bugeon et al. (2022)) that belonged to a subtype with at least 3 cells. We used the previous assignment into 35 upper-layer subtypes and grouped Serpinf1 cells into the Vip class. Consistent with the original publication, this resulted in 1065 cells, hierarchically distributed across 5 subclasses (e.g., Pvalb) and 35 subtypes (e.g.,

Pvalb-Reln-Itm2a). The Npy gene count of 58 cells was missing (NaN); we assumed these values were missing at random and imputed them with the subtype-specific median value. Zero-imputation gave similar results. We computed each cell's average activity per behavioural state. Whenever a cell was recorded during multiple sessions, we used the session with the longest period of "stationary synchronised" activity since this was the least frequent state. Since 193 cells were not recorded during the stationary synchronised state, state modulation was computed for the remaining 872 cells. The expression matrix contained continuously valued estimates of gene expression instead of integer counts. We normalized these values to 10,000 "counts" per cell for consistency with the other datasets, although this slightly decreased predictive performance. Finally, we log-transformed the normalized values after adding one pseudo-count, $\log(1 + x)$. The log transformation is a widely used preprocessing step in the analysis of count data (Luecken and Theis, 2019; Ahlmann-Eltze and Huber, 2023), although other transformations are also possible (see, e.g., Hafemeister and Satija (2019) and Lause et al. (2021)). Linear least squares regression was used to predict state modulation from individual PCs or cholinergic receptors; cross-validated ridge regression was used to predict state modulation from multiple PCs to mitigate overfitting.

Other datasets

The transcriptomic datasets each consisted of raw count matrices and metadata that included cell class and subtype/cluster. For the Tasic dataset, we only considered the VISp (not the ALM) cells to allow for a direct comparison with Bugeon et al. In the mouse datasets, we merged the small number of Serpinf1 cells into the Vip cluster for consistency with the analyses from ref. (Bugeon et al., 2022). From the Colquitt and Tosches datasets, we only used the zebra finch and turtle cells, respectively, since the data from other species (Bengalese finch and lizard) contained only a small number of interneurons. For both datasets, we assigned cells to putative mammalian homologues according to the correlation-based matching in the original publications (Tosches et al., 2018; Colquitt et al., 2021). For the Hodge dataset, we assigned each cell a cell type based on the original publication (Hodge et al., 2019). Genes were named according to the mouse convention (e.g., "Npy" instead of the human "NPY"). We selected the 11968 shared genes by intersecting the gene lists from all datasets (except the Bugeon data, which has 72 genes) to put the datasets in a shared space.

Principal component analysis (PCA)

We scaled gene expression values to 10,000 counts per cell (CP10K) to account for differences in sequencing depth across cells and log-transformed the normalized data. We then identified the top 2000 highly variable genes based on their dispersion across cells (Scanpy's `highly_variable_genes`; using 3000 genes gave similar results). We computed the top 30 PCs based on these highly variable genes. For visual comparison, we made an arbitrary but consistent choice for the signs of tPC1 and tPC2.

We quantified the similarity of PCs from different datasets using principal angles (Björck and Golub (1973); Scipy's `subspace_angles`). More precisely, let W_X be the gene-by-PC matrix whose columns are the PCs of dataset A. The principal angles between the PC subspaces of datasets A and B are computed from the singular value decomposition (SVD) of

the PC-by-PC matrix $W_A^T W_B$, i.e.

$$W_A^T W_B = U \Sigma V^T.$$

The columns of U and V contain paired linear combinations of PCs from datasets A and B, respectively, ordered by principal angles. The diagonal matrix Σ contains the singular values σ_i. The ith principal angle from the corresponding singular value σ_i is computed as

$$\vartheta_i = \arccos(\sigma_i).$$

As a complementary measure of PC subspace similarity, we computed the variance explained in one dataset by the top PCs of another dataset. Let $\mathbf{w}_{A,i}$ be the ith PC of dataset A, and let C^B be the covariance matrix of dataset B. The ith PC of dataset A explains an amount of variance in dataset B equal to

$$\mathbf{w}_{A,i}^T C^B \mathbf{w}_{A,i}.$$

For each pairwise comparison, we computed the covariance and PCs only from genes that were highly variable in both datasets. This was done to avoid the computation of large covariance matrices. For comparison, the variance of each PC was normalized by the variance explained by the first PC of the original dataset:

$$\text{normalized variance} = \frac{\mathbf{w}_{B,i}^T C^A \mathbf{w}_{B,i}}{\mathbf{w}_{A,1}^T C^A \mathbf{w}_{A,1}}.$$

Chance level was estimated by computing the variance explained by a random, normalized vector. To predict state modulation from the tPCs of other species, we first intersected their gene sets with the 72 genes from Bugeon et al. (Bugeon et al., 2022) We then separately preprocessed both datasets. Finally, we projected the Bugeon data onto tPC1 from the secondary dataset and used this to predict state modulation. Performance was quantified using leave-one-out R^2 and the Pearson correlation coefficient.

Subsampling gene counts

The datasets vary in their sequencing depth (the number of RNA counts per cell, see Table 6.1), presumably due to a combination of technical and biological differences. We aimed to control for these differences by downsampling counts to the depth of the shallower dataset, as follows. Let X_{cg}^d be the number of counts from gene g in cell c of dataset d. We defined the count depth of a dataset as the average counts per cell:

$$\text{depth}(d) = \frac{1}{C \cdot G} \sum_{c,g} X_{c,g}^d.$$

Here C is the number of cells, and G is the number of genes. If d_1 is the shallowest dataset, and d_2 is a deeper-sequenced dataset, we define their relative sequencing depth as

$$p = \frac{\text{depth}(d_1)}{\text{depth}(d_2)} \in (0, 1).$$

To match the sequencing depth of the shallower dataset, we keep each gene count with a probability p:

$$\hat{X}^{d_2}_{c,g} \sim \text{Binomial}(X^{d_2}_{c,g}, p).$$

This subsampling procedure equalizes the sequencing depth of the down-sampled dataset to that of the shallower dataset. The Tasic and Hodge data served as the reference datasets for comparison with the mouse and human expression values, respectively, since these were the deepest datasets.

Dataset integration

We used Seurat's anchor-based integration (Stuart et al., 2019) to map datasets onto the Tasic data (Fig. B.5). To this end, we converted the AnnData objects to Seurat objects. Next, we separately log normalized each dataset as described above (this time using the equivalent Seurat function `NormalizeData`), and found genes that were highly variable across datasets (`FindVariableFeatures`, followed by `SelectIntegrationFeatures`, with 2000 features). Next, we found mutual nearest neighbours across datasets ("anchors") after projecting each dataset onto the other's PCA space (reciprocal PCA). A more flexible reduction method (canonical correlation analysis) gave similar results. We then used the anchors to identify and project out dataset-specific differences. After integration, PCA was performed separately on each transformed dataset.

Network simulations

We simulated a rate-based network of Pvalb, Sst, and Vip interneurons and excitatory pyramidal neurons. A single equation represented each cell type except for the pyramidal neurons, represented by two equations, modelling the somatic and dendritic compartments. The network state was defined by the rate vector $\mathbf{r} = (r_e, r_d, r_p, r_s, r_v)$, of somatic, dendritic, Pvalb, Sst, and Vip activity. The rate of cell type/compartment x evolved according to

$$\tau_x \dot{u}_x = -u_x + \sum_{y \neq x} w_{x,y} r_y + I_{x,0} + I_x + m_x, \quad x \in \{e, d, p, s, v\}, \tag{6.1}$$

$$r_x = f(u_x). \tag{6.2}$$

Here τ_x is the membrane time constant (2 ms for excitatory cells, 10 ms for inhibitory cells), $f(u) = \max(u, 0)$ is the rectified linear activation function, and the w_{xy} are recurrent weights. $I_{x,0}$ is a constant background input that sets the baseline rate, I_x is a time-varying external input, and m_x is an additive cholinergic modulation. We will refer to m_x as a cell's cholinergic receptor density to distinguish it from the "effective" cholinergic modulation, which also depends on the network dynamics (see Network analysis).

The recurrent connections were chosen based on experimental (Pfeffer et al., 2013; Campagnola et al., 2022) and theoretical work (Hertäg and Sprekeler, 2019). The only difference is relatively weak mutual inhibition between Sst and Vip neurons; strong inhibition could prevent the simultaneous activation of these cell types observed in the data (Bugeon et al., 2022).

$$W = \begin{pmatrix} w_{ee} & w_{ed} & w_{ep} & w_{es} & w_{ev} \\ w_{de} & w_{dd} & w_{dp} & w_{ds} & w_{dv} \\ w_{pe} & w_{pd} & w_{pp} & w_{ps} & w_{pv} \\ w_{se} & w_{sd} & w_{sp} & w_{ss} & w_{sv} \\ w_{ve} & w_{vd} & w_{vp} & w_{vs} & w_{vv} \end{pmatrix} = \begin{pmatrix} 0.42 & 1 & -0.42 & 0 & 0 \\ 0.042 & 0 & 0 & -0.49 & 0 \\ 0.45 & 0 & -0.75 & -0.78 & 0 \\ 0.35 & 0 & 0 & 0 & -0.175 \\ 1 & 0 & 0 & -0.175 & 0 \end{pmatrix} \tag{6.3}$$

The background inputs $I_{x,0}$ were set to achieve the following baseline rates:

$$(r_e, r_d, r_p, r_s, r_v) = (1, 1, 8, 4, 3) \quad (1/\text{s}).$$

The external inputs to pyramidal soma and dendrites were defined as:

$$I_x(t) = 1 + .5 \sin(\varphi_x t), \quad x \in \{e, d\}.$$

with $\varphi_e = 1/300$ ms (soma) or $\varphi_d = 1/70$ ms (dendrite).

The interneurons received cholinergic modulation instead of external inputs, and their amplitudes were varied based on qualitative differences in cholinergic receptor expression (Table 6.3). These amplitudes were the only differences between species-specific networks. In the mouse network, Pvalb neurons were negatively modulated; Vip and — to a lesser extent — Sst neurons were positively modulated. This is consistent with both the activity and expression data from the mouse. For the other species, we only have expression data. In the human and turtle network, Pvalb neurons were not modulated—consistent with their weak or absent expression of, e.g., Chrm4. In the turtle network, Sst neurons were negatively modulated; Vip neurons were positively modulated, but to a smaller extent, given the under-expression of Chrm3 and Chrna5 in the turtle data. For the turtle network, we added a positive external input (amplitude 5) to the Sst equation in the absence of cholinergic modulation. A similar result could be obtained by decreasing the dendritic drive during baseline.

species \ cell type	Pvalb	Sst	Vip
Mouse	-2	6	8
Human	0	6	8
Turtle	0	-1	4

Table 6.3: Network parameters: species and cell type-specific additive cholinergic modulation

The network dynamics were numerically integrated using a forward Euler scheme with a time step of 0.1 milliseconds. Each simulation consisted of 11000 time steps divided into a baseline period of 3300 steps, a cholinergic modulation time of 3400 steps, and another baseline period of 4300 steps. Not shown in Fig. 6.5 is an initial settling time of 300 timesteps. These values were chosen to let the figure highlight the effect of turning the modulation on and off.

7 Discussion

The aim of this thesis was to study the function and, coincidentally, the origins of interneuron diversity. We found that certain aspects of interneuron diversity can be explained by extending the function of excitation/inhibition balance to multiple cellular compartments. We also discovered, however, that this current function might not be the reason why interneuron diversity originally evolved. Finally, we found that although interneurons of different species share an evolutionary history, they have evolved their own patterns of gene expression that are likely functionally relevant.

This work used two relatively new methods: gradient-based optimization and single-cell transcriptomics. These methods, while highly influential in contemporary neuroscience, embody very different approaches to studying the brain and operate at distinct levels of analysis. The optimization-based approach is functional and abstract; transcriptomics is descriptive and reductionistic. We will end with a discussion of important challenges for integrating functional and descriptive views of the brain.

7.1 Degeneracy complicates the use of task-optimized models

The same biological function can often be performed by vastly different structures, a phenomenon known as degeneracy (Mason et al., 2015; Goaillard and Marder, 2021). In this thesis, we encountered redundancy twice: a compartment-specific E/I balance could be accomplished by discrete interneuron types but also by a continuum (Ch. 4), and similar state modulation resulted from different cholinergic receptor distributions (Ch. 6).

Redundancy implies a many-to-one function from structure to function; in other words: function does not determine structure. This clearly presents a challenge for models designed to explain structure from function, as we aimed to do for interneuron circuits in Ch. 4. Facing an underdetermined task, most approaches for deriving such task-optimized models will find a single solution, leaving the modeller unaware of alternatives. If, however, biological data is available for comparison, the task or network architecture can be adapted until the result matches the data (e.g., Sussillo et al. (2015)). An ostensible failure has been turned into a

potential success by revealing unanticipated constraints. Of course, the biological relevance of these constraints has to be tested.

In Ch. 4, mathematical analyses revealed why our optimisation created a continuum of interneuron types: inhibitory connectivity could cancel, for example, somatic inputs to dendrite-projecting neurons, creating selective feedback from unselective activity. Cancellation is not possible when each compartment experiences periods of inactivity. This requirement for a biologically realistic solution is not implausible, but it is unlikely the only reason why PV+ and SST+ interneurons are discrete.

A more fundamental reason is found in the developmental mechanisms that generate PV+ and SST+ interneurons (Ch. 5). For example, dendrite targeting neurons that migrate via a different route that allows them to target dendrites (Lim et al., 2018b). More generally, broad cell types form discrete entities through mutually repressive transcriptional programs (Greig et al., 2013). Discrete cell types therefore result from genetic constraints and the generative rules of development rather than the external environment.

7.2 Adaptationism *and* structuralism

The competing explanations of interneuron diversity —function vs. constraint— reflect a long-standing debate in evolutionary biology between viewpoints known as adaptationism and structuralism. *Adaptationism* interprets biological structure as the result of adaptation to selective pressure. This approach has also been referred to as *externalism* for its emphasis on the physical environment to determine selective pressures (Alberch, 1989). In the words of geneticist Dobzhansky: "The enormous diversity of organisms may be envisaged as correlated with the immense variety of environments and of ecological niches which exist on earth." (Dobzhansky, 1982).

Structuralism or *internalism* complements the adaptationist perspective by emphasising the role of constraints and generative rules in explaining biological order (Alberch, 1989). Perhaps its starkest illustration is the many genetic deficits that lead to highly structured rather than "random" malformations. For example, misexpression of the Pax6 protein during *Drosophila* development can cause an entire eye to form in an abnormal site, such as a wing or leg (Halder et al., 1995). A less dramatic but directly relevant example: the misexpression of cell-type specific transcription factors can lead a cell to adopt the fate of a different, but equally 'valid', cell type instead of some intermediate phenotype (e.g., (Lodato et al., 2014)).

The point is not to debate the relative importance of selection versus constraints: clearly, both play a role and the two interact (Uller et al., 2018). The point is that although multiple processes play a role, adaptation receives virtually all the attention in computational neuroscience. Confronted with new data, many computational neuroscientists will instinctively ask what purpose the observed feature serves. The implicit assumption is that the existence and specific manifestation of any trait must serve an adaptive, preferably computational, function. The failure of one adaptive explanation immediately leads to the search for another; the lack of optimality in one domain must imply a trade-off with competing goals.

Precisely this line of reasoning was criticised by Gould and Lewontin in their influential 1979 "Spandrels" paper (Gould and Lewontin (1979); Fig. 7.1). Specifically, Gould and

Lewontin criticised the tendency of evolutionary biologists to assume rather than demonstrate the role of natural selection. However, selection is biased by genetic and developmental constraints (Brakefield, 2006), and many biological features were not selected for their current function (Gould and Vrba, 1982; Conant and Wolfe, 2008). Simply put: evolution is not the same as adaptation or natural selection. Blindly assuming that it is prematurely rules out promising valid hypotheses.

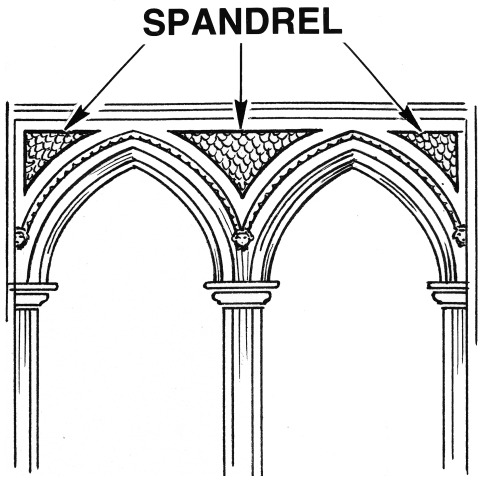

Figure 7.1: Spandrels: a metaphor for for adaptive storytelling. Spandrels, the curved areas inevitably left between cathedral arches, are a famous metaphor for an overreliance on adaptive hypotheses. Although the result of an architectural constraint, spandrels are often decorated with artworks "so elaborate, harmonious, and purposeful that we are tempted to view it as the starting point of any analysis, as the cause in some sense of the surrounding architecture (Gould and Lewontin, 1979). Taking their cues from neuroscience colleagues, theoretical architects might even show that spandrels: maximize information, synchronize distant parts of the cathedral, emerge during the optimization of an artificial cathedral, represent a Bayesian prior over biblical history, and minimize free energy. This does not, however, imply that these functions are the reason that spandrels exist. Figure from `commons.wikimedia.org`, released into the public domain.

Since the Sprandrels critique, evolutionary biologists have become more careful in producing just-so stories (Pigliucci and Kaplan, 2000), using statistics (Lande and Arnold, 1983; McDonald and Kreitman, 1991) and experiments (Dean and Thornton, 2007; Barrett and Hoekstra, 2011) to test for natural selection. Neuroscientists might similarly benefit from a nuanced and empirically-founded understanding of evolution and development. This could be especially important for the study of circuits and cell types: whereas abstract population activity is reasonably constrained by a particular task, its implementation will partially depend on evolutionary history and developmental constraints.

7.3 Whence interneuron diversity?

The importance of evolutionary and developmental constraints does not render optimization-based models useless for the study of structure-function relationships. In this thesis, for example, optimization revealed which features of interneuron circuits are necessary for precise inhibition (Ch. 4). Similar approaches have recently shown that optimization can create heterogeneous time constants (Perez-Nieves et al., 2021), biologically-relevant transfer functions (Geadah et al., 2022) and spatially-localized connectivity (Schaeffer et al., 2023; Achterberg et al., 2023). But such purely functional approach are unlikely to explain the existence and properties of

specific cell types. Instead, this will require the careful combination of quantitative models with detailed experiments.

References

(N.d.). https://creativecommons.org/licenses/by-nc-sa/3.0/.

(N.d.). https://creativecommons.org/licenses/by/4.0/.

Abs, E., R. B. Poorthuis, D. Apelblat, K. Muhammad, M. B. Pardi, L. Enke, D. Kushinsky, D.-L. Pu, M. F. Eizinger, K.-K. Conzelmann, et al. (2018). "Learning-related plasticity in dendrite-targeting layer 1 interneurons". In: *Neuron* 100.3, pp. 684–699.

Achterberg, J., D. Akarca, D. Strouse, J. Duncan, and D. E. Astle (2023). "Spatially embedded recurrent neural networks reveal widespread links between structural and functional neuroscience findings". In: *Nature Machine Intelligence*, pp. 1–13.

Adler, A., R. Zhao, M. E. Shin, R. Yasuda, and W.-B. Gan (2019). "Somatostatin-expressing interneurons enable and maintain learning-dependent sequential activation of pyramidal neurons". In: *Neuron* 102.1, pp. 202–216.

Ahlmann-Eltze, C. and W. Huber (2023). "Comparison of transformations for single-cell RNA-seq data". In: *Nature Methods*, pp. 1–8.

Alberch, P. (1989). "The logic of monsters: evidence for internal constraint in development and evolution". In: *Geobios* 22, pp. 21–57.

Alcamo, E. A., L. Chirivella, M. Dautzenberg, G. Dobreva, I. Fariñas, R. Grosschedl, and S. K. McConnell (2008). "Satb2 regulates callosal projection neuron identity in the developing cerebral cortex". In: *Neuron* 57.3, pp. 364–377.

Aldridge, S. and S. A. Teichmann (2020). "Single cell transcriptomics comes of age". In: *Nature Communications* 11.1, p. 4307.

Alexander, R. M. (1996). *Optima for animals*. Princeton university press.

Alitto, H. J. and Y. Dan (2013). "Cell-type-specific modulation of neocortical activity by basal forebrain input". In: *Frontiers in systems neuroscience* 6, p. 79.

Alon, U. (2019). *An introduction to systems biology: design principles of biological circuits*. CRC press.

Alser, M., J. Rotman, D. Deshpande, K. Taraszka, H. Shi, P. I. Baykal, H. T. Yang, V. Xue, S. Knyazev, B. D. Singer, et al. (2021). "Technology dictates algorithms: recent developments in read alignment". In: *Genome biology* 22.1, p. 249.

Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman (1990). "Basic local alignment search tool". In: *Journal of molecular biology* 215.3, pp. 403–410.

Anaconda Software Distribution (2020). Version Vers. 2-2.4.0.

Anderson, S., D. Eisenstat, L. Shi, and J. Rubenstein (1997). "Interneuron migration from basal forebrain to neocortex: dependence on Dlx genes". In: *Science* 278.5337, pp. 474–476.

Andrews, T. S., V. Y. Kiselev, D. McCarthy, and M. Hemberg (2021). "Tutorial: guidelines for the computational analysis of single-cell RNA sequencing data". In: *Nature protocols* 16.1, pp. 1–9.

Arendt, D., J. M. Musser, C. V. Baker, A. Bergman, C. Cepko, D. H. Erwin, M. Pavlicev, G. Schlosser, S. Widder, M. D. Laubichler, et al. (2016). "The origin and evolution of cell types". In: *Nature Reviews Genetics* 17.12, pp. 744–757.

Avendano, C., D. Umbriaco, R. Dykes, and L. Descarries (1996). "Acetylcholine innervation of sensory and motor neocortical areas in adult cat: a choline acetyltransferase immunohistochemical study". In: *Journal of chemical neuroanatomy* 11.2, pp. 113–130.

Bakken, T. E., R. D. Hodge, J. A. Miller, Z. Yao, T. N. Nguyen, B. Aevermann, E. Barkan, D. Bertagnolli, T. Casper, N. Dee, et al. (2018). "Single-nucleus and single-cell transcriptomes compared in matched cortical cell types". In: *PloS one* 13.12, e0209648.

Bakken, T. E., N. L. Jorstad, Q. Hu, B. B. Lake, W. Tian, B. E. Kalmbach, M. Crow, R. D. Hodge, F. M. Krienen, S. A. Sorensen, et al. (2021). "Comparative cellular analysis of motor cortex in human, marmoset and mouse". In: *Nature* 598.7879, pp. 111–119.

Bandler, R. C., I. Vitali, R. N. Delgado, M. C. Ho, E. Dvoretskova, J. S. Ibarra Molinas, P. W. Frazel, M. Mohammadkhani, R. Machold, S. Maedler, et al. (2021). "Single-cell delineation of lineage and genetic identity in the mouse brain". In: *Nature*, pp. 1–6.

Banino, A., C. Barry, B. Uria, C. Blundell, T. Lillicrap, P. Mirowski, A. Pritzel, M. J. Chadwick, T. Degris, J. Modayil, et al. (2018). "Vector-based navigation using grid-like representations in artificial agents". In: *Nature* 557.7705, pp. 429–433.

Barak, O. (2017). "Recurrent neural networks as versatile tools of neuroscience research". In: *Current opinion in neurobiology* 46, pp. 1–6.

Barak, O., D. Sussillo, R. Romo, M. Tsodyks, and L. Abbott (2013). "From fixed points to chaos: three models of delayed discrimination". In: *Progress in neurobiology* 103, pp. 214–222.

Barrett, R. D. and H. E. Hoekstra (2011). "Molecular spandrels: tests of adaptation at the genetic level". In: *Nature Reviews Genetics* 12.11, pp. 767–780.

Bartos, M., I. Vida, M. Frotscher, A. Meyer, H. Monyer, J. R. Geiger, and P. Jonas (2002). "Fast synaptic inhibition promotes synchronized gamma oscillations in hippocampal interneuron networks". In: *Proceedings of the National Academy of Sciences* 99.20, pp. 13222–13227.

Bathellier, B., T. W. Margrie, and M. E. Larkum (2009). "Properties of piriform cortex pyramidal cell dendrites: implications for olfactory circuit design". In: *Journal of Neuroscience* 29.40, pp. 12641–12652.

Batista-Brito, R., E. Zagha, J. M. Ratliff, and M. Vinck (2018). "Modulation of cortical circuits by top-down processing and arousal state in health and disease". In: *Current opinion in neurobiology* 52, pp. 172–181.

Baydin, A. G., B. A. Pearlmutter, A. A. Radul, and J. M. Siskind (2018). "Automatic differentiation in machine learning: a survey". In: *Journal of Marchine Learning Research* 18, pp. 1–43.

Beaulieu-Laroche, L., N. J. Brown, M. Hansen, E. H. Toloza, J. Sharma, Z. M. Williams, M. P. Frosch, G. R. Cosgrove, S. S. Cash, and M. T. Harnett (2021). "Allometric rules for mammalian cortical layer 5 neuron biophysics". In: *Nature*, pp. 1–5.

Beaulieu-Laroche, L., E. H. Toloza, N. J. Brown, and M. T. Harnett (2019). "Widespread and highly correlated somato-dendritic activity in cortical layer 5 neurons". In: *Neuron* 103.2, pp. 235–241.

Beierlein, M. and B. W. Connors (2002). "Short-term dynamics of thalamocortical and intracortical synapses onto layer 6 neurons in neocortex". In: *Journal of neurophysiology* 88.4, pp. 1924–1932.

Beiran, M., N. Meirhaeghe, H. Sohn, M. Jazayeri, and S. Ostojic (2023). "Parametric control of flexible timing through low-dimensional neural manifolds". In: *Neuron* 111.5, pp. 739–753.

Bellec, G., D. Salaj, A. Subramoney, R. Legenstein, and W. Maass (2018). "Long short-term memory and learning-to-learn in networks of spiking neurons". In: *arXiv preprint arXiv:1803.09574*.

Bellec, G., F. Scherr, A. Subramoney, E. Hajek, D. Salaj, R. Legenstein, and W. Maass (2020). "A solution to the learning dilemma for recurrent networks of spiking neurons". In: *Nature communications* 11.1, p. 3625.

Benezra, S. E., R. T. Narayanan, R. Egger, M. Oberlaender, and M. A. Long (2018). "Morphological characterization of HVC projection neurons in the zebra finch (Taeniopygia guttata)". In: *Journal of Comparative Neurology* 526.10, pp. 1673–1689.

Beniaguev, D., I. Segev, and M. London (2021). "Single cortical neurons as deep artificial neural networks". In: *Neuron* 109.17, pp. 2727–2739.

Berg, J., S. A. Sorensen, J. T. Ting, J. A. Miller, T. Chartrand, A. Buchin, T. E. Bakken, A. Budzillo, N. Dee, S.-L. Ding, et al. (2021). "Human neocortical expansion involves glutamatergic neuron diversification". In: *Nature* 598.7879, pp. 151–158.

Berger, T. K., G. Silberberg, R. Perin, and H. Markram (2010). "Brief bursts self-inhibit and correlate the pyramidal network". In: *PLoS biology* 8.9, e1000473.

Bernaerts, Y., M. Deistler, P. J. Goncalves, J. Beck, M. Stimberg, F. Scala, A. S. Tolias, J. H. Macke, D. Kobak, and P. Berens (2023). "Combined statistical-mechanistic modeling links ion channel genes to physiology of cortical neuron types". In: *bioRxiv*, pp. 2023–03.

Bhatia, A., S. Moza, and U. S. Bhalla (2019). "Precise excitation-inhibition balance controls gain and timing in the hippocampus". In: *Elife* 8, e43415.

Bialek, W. (2012). *Biophysics: searching for principles*. Princeton University Press.

Björck, A. and G. H. Golub (1973). "Numerical methods for computing angles between linear subspaces". In: *Mathematics of computation* 27.123, pp. 579–594.

Blanton, M. G., J. M. Shen, and A. R. Kriegstein (1987). "Evidence for the inhibitory neurotransmitter γ-aminobutyric acid in aspiny and sparsely spiny nonpyramidal neurons of the turtle dorsal cortex". In: *Journal of Comparative Neurology* 259.2, pp. 277–297.

Boldog, E., T. E. Bakken, R. D. Hodge, M. Novotny, B. D. Aevermann, J. Baka, S. Bordé, J. L. Close, F. Diez-Fuertes, S.-L. Ding, et al. (2018). "Transcriptomic and morphophysiological evidence for a specialized human cortical GABAergic cell type". In: *Nature neuroscience* 21.9, pp. 1185–1195.

Bomkamp, C., S. J. Tripathy, C. Bengtsson Gonzales, J. Hjerling-Leffler, A. M. Craig, and P. Pavlidis (2019). "Transcriptomic correlates of electrophysiological and morphological diversity within and across excitatory and inhibitory neuron classes". In: *PLoS computational biology* 15.6, e1007113.

Bradbury, J., R. Frostig, P. Hawkins, M. J. Johnson, C. Leary, D. Maclaurin, G. Necula, A. Paszke, J. VanderPlas, S. Wanderman-Milne, et al. (2018). "JAX: composable transformations of Python+ NumPy programs". In.

Brakefield, P. M. (2006). "Evo-devo and constraints on selection". In: *Trends in Ecology & Evolution* 21.7, pp. 362–368.

Bratsch-Prince, J. X., J. W. Warren III, G. C. Jones, A. J. McDonald, and D. D. Mott (2023). "Acetylcholine engages distinct amygdala microcircuits to gate internal theta rhythm". In: *bioRxiv*, pp. 2023–02.

Brunel, N. (2000). "Dynamics of sparsely connected networks of excitatory and inhibitory spiking neurons". In: *Journal of computational neuroscience* 8.3, pp. 183–208.

Brunel, N. and M. C. Van Rossum (2007). "Lapicque's 1907 paper: from frogs to integrate-and-fire". In: *Biological cybernetics* 97.5-6, pp. 337–339.

Bugeon, S., J. Duffield, M. Dipoppa, A. Ritoux, I. Prankerd, D. Nicoloutsopoulos, D. Orme, M. Shinn, H. Peng, H. Forrest, et al. (2022). "A transcriptomic axis predicts state modulation of cortical interneurons". In: *Nature* 607.7918, pp. 330–338.

Burgess, D. J. (2019). "Spatial transcriptomics coming of age". In: *Nature Reviews Genetics* 20.6, pp. 317–317.

Butt, S. J., M. Fuccillo, S. Nery, S. Noctor, A. Kriegstein, J. G. Corbin, and G. Fishell (2005). "The temporal and spatial origins of cortical interneurons predict their physiological subtype". In: *Neuron* 48.4, pp. 591–604.

Butt, S. J., V. H. Sousa, M. V. Fuccillo, J. Hjerling-Leffler, G. Miyoshi, S. Kimura, and G. Fishell (2008). "The requirement of Nkx2-1 in the temporal specification of cortical interneuron subtypes". In: *Neuron* 59.5, pp. 722–732.

Cadwell, C. R., A. Palasantza, X. Jiang, P. Berens, Q. Deng, M. Yilmaz, J. Reimer, S. Shen, M. Bethge, K. F. Tolias, et al. (2016). "Electrophysiological, transcriptomic and morphologic profiling of single neurons using Patch-seq". In: *Nature biotechnology* 34.2, pp. 199–203.

Caillard, O., H. Moreno, B. Schwaller, I. Llano, M. R. Celio, and A. Marty (2000). "Role of the calcium-binding protein parvalbumin in short-term synaptic plasticity". In: *Proceedings of the National Academy of Sciences* 97.24, pp. 13372–13377.

Campagnola, L., S. C. Seeman, T. Chartrand, L. Kim, A. Hoggarth, C. Gamlin, S. Ito, J. Trinh, P. Davoudian, C. Radaelli, et al. (2022). "Local connectivity and synaptic dynamics in mouse and human neocortex". In: *Science* 375.6585, eabj5861.

Cardin, J. A. (2019). "Functional flexibility in cortical circuits". In: *Current opinion in neurobiology* 58, pp. 175–180.

Chartrand, T., R. Dalley, J. Close, N. A. Goriounova, B. R. Lee, R. Mann, J. A. Miller, G. Molnar, A. Mukora, L. Alfiler, et al. (2023). "Morphoelectric and transcriptomic divergence of the layer 1 interneuron repertoire in human versus mouse neocortex". In: *Science* 382.6667, eadf0805.

Chen, K. H., A. N. Boettiger, J. R. Moffitt, S. Wang, and X. Zhuang (2015a). "Spatially resolved, highly multiplexed RNA profiling in single cells". In: *Science* 348.6233, aaa6090.

Chen, N., H. Sugihara, and M. Sur (2015b). "An acetylcholine-activated microcircuit drives temporal dynamics of cortical activity". In: *Nature neuroscience* 18.6, pp. 892–902.

Chen, X., S. Fischer, M. C. Rue, A. Zhang, D. Mukherjee, P. O. Kanold, J. Gillis, and A. M. Zador (2022). "Whole-cortex in situ sequencing reveals peripheral input-dependent cell type-defined area identity". In: *BioRxiv*, pp. 2022–11.

Chen, X., Y.-C. Sun, H. Zhan, J. M. Kebschull, S. Fischer, K. Matho, Z. J. Huang, J. Gillis, and A. M. Zador (2019). "High-throughput mapping of long-range neuronal projection using in situ sequencing". In: *Cell* 179.3, pp. 772–786.

Chiu, C. Q., G. Lur, T. M. Morse, N. T. Carnevale, G. C. Ellis-Davies, and M. J. Higley (2013). "Compartmentalization of GABAergic inhibition by dendritic spines". In: *Science* 340.6133, pp. 759–762.

Cho, K., B. Van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio (2014). "Learning phrase representations using RNN encoder-decoder for statistical machine translation". In: *arXiv preprint arXiv:1406.1078*.

Chung, S. and L. Abbott (2021). "Neural population geometry: An approach for understanding biological and artificial neural networks". In: *Current opinion in neurobiology* 70, pp. 137–144.

Colquitt, B. M., D. P. Merullo, G. Konopka, T. F. Roberts, and M. S. Brainard (2021). "Cellular transcriptomics reveals evolutionary identities of songbird vocal circuits". In: *Science* 371.6530, eabd9704.

Conant, G. C. and K. H. Wolfe (2008). "Turning a hobby into a job: how duplicated genes find new functions". In: *Nature Reviews Genetics* 9.12, pp. 938–950.

Condylis, C., A. Ghanbari, N. Manjrekar, K. Bistrong, S. Yao, Z. Yao, T. N. Nguyen, H. Zeng, B. Tasic, and J. L. Chen (2022). "Dense functional and molecular readout of a circuit hub in sensory cortex". In: *Science* 375.6576, eabl5981.

Connors, B. W. and A. Kriegstein (1986). "Cellular physiology of the turtle visual cortex: distinctive properties of pyramidal and stellate neurons". In: *Journal of Neuroscience* 6.1, pp. 164–177.

Coppola, J. J. and A. A. Disney (2018). "Is there a canonical cortical circuit for the cholinergic system? Anatomical differences across common model systems". In: *Frontiers in neural circuits* 12, p. 8.

Costa, R. P., P. J. Sjöström, and M. C. Van Rossum (2013). "Probabilistic inference of short-term synaptic plasticity in neocortical microcircuits". In: *Frontiers in computational neuroscience* 7, p. 75.

Cramer, B., Y. Stradmann, J. Schemmel, and F. Zenke (2020). "The heidelberg spiking data sets for the systematic evaluation of spiking neural networks". In: *IEEE Transactions on Neural Networks and Learning Systems* 33.7, pp. 2744–2757.

Cristo, G. D., C. Wu, B. Chattopadhyaya, F. Ango, G. Knott, E. Welker, K. Svoboda, and Z. J. Huang (2004). "Subcellular domain-restricted GABAergic innervation in primary visual cortex in the absence of sensory and thalamic inputs". In: *Nature neuroscience* 7.11, pp. 1184–1186.

Cueva, C. J. and X.-X. Wei (2018). "Emergence of grid-like representations by training recurrent neural networks to perform spatial localization". In: *arXiv preprint arXiv:1803.07770*.

Dard, R. F., E. Leprince, J. Denis, S. R. Balappa, D. Suchkov, R. Boyce, C. Lopez, M. Giorgi-Kurz, T. Szwagier, T. Dumont, et al. (2022). "The rapid developmental rise of somatic inhibition disengages hippocampal dynamics from self-motion". In: *Elife* 11, e78116.

Dean, A. M. and J. W. Thornton (2007). "Mechanistic approaches to the study of evolution: the functional synthesis". In: *Nature Reviews Genetics* 8.9, pp. 675–688.

Dehorter, N., G. Ciceri, G. Bartolini, L. Lim, I. Del Pino, and O. Marín (2015). "Tuning of fast-spiking interneuron properties by an activity-dependent transcriptional switch". In: *Science* 349.6253, pp. 1216–1220.

Denaxa, M., G. Neves, A. Rabinowitz, S. Kemlo, P. Liodis, J. Burrone, and V. Pachnis (2018). "Modulation of apoptosis controls inhibitory interneuron number in the cortex". In: *Cell reports* 22.7, pp. 1710–1721.

Devoogd, T. J. and F. Nottebohm (1981). "Sex differences in dendritic morphology of a song control nucleus in the canary: a quantitative Golgi study". In: *Journal of Comparative Neurology* 196.2, pp. 309–316.

Dipoppa, M., A. Ranson, M. Krumin, M. Pachitariu, M. Carandini, and K. D. Harris (2018). "Vision and locomotion shape the interactions between neuron types in mouse visual cortex". In: *Neuron* 98.3, pp. 602–615.

Disney, A. A., C. Aoki, and M. J. Hawken (2007). "Gain modulation by nicotine in macaque v1". In: *Neuron* 56.4, pp. 701–713.

Disney, A. A. and J. H. Reynolds (2014). "Expression of m1-type muscarinic acetylcholine receptors by parvalbumin-immunoreactive neurons in the primary visual cortex: A comparative study of rat, guinea pig, ferret, macaque, and human". In: *Journal of Comparative Neurology* 522.5, pp. 986–1003.

Dobzhansky, T. (1982). *Genetics and the Origin of Species*. 11. Columbia university press.

Dolan, J. and K. J. Mitchell (2013). "Mutation of Elfn1 in mice causes seizures and hyperactivity". In: *PloS one* 8.11, e80491.

Dolan, J., K. Walshe, S. Alsbury, K. Hokamp, S. O'Keeffe, T. Okafuji, S. F. Miller, G. Tear, and K. J. Mitchell (2007). "The extracellular leucine-rich repeat superfamily; a comparative survey and analysis of evolutionary relationships and expression patterns". In: *BMC genomics* 8.1, pp. 1–24.

Driscoll, L., K. Shenoy, and D. Sussillo (2022). "Flexible multitask computation in recurrent networks utilizes shared dynamical motifs". In: *bioRxiv*, pp. 2022–08.

Dudok, B., P. M. Klein, E. Hwaun, B. R. Lee, Z. Yao, O. Fong, J. C. Bowler, S. Terada, F. T. Sparks, G. G. Szabo, et al. (2021). "Alternating sources of perisomatic inhibition during behavior". In: *Neuron* 109.6, pp. 997–1012.

Dunn, H. A., D. N. Patil, Y. Cao, C. Orlandi, and K. A. Martemyanov (2018). "Synaptic adhesion protein ELFN1 is a selective allosteric modulator of group III metabotropic glutamate receptors in trans". In: *Proceedings of the National Academy of Sciences* 115.19, pp. 5022–5027.

Edgar, R. C. (2022). "Muscle5: High-accuracy alignment ensembles enable unbiased assessments of sequence homology and phylogeny". In: *Nature Communications* 13.1, p. 6968.

Eng, C.-H. L., M. Lawson, Q. Zhu, R. Dries, N. Koulena, Y. Takei, J. Yun, C. Cronin, C. Karp, G.-C. Yuan, et al. (2019). "Transcriptome-scale super-resolved imaging in tissues by RNA seqFISH+". In: *Nature* 568.7751, pp. 235–239.

Estebanez, L., D. Hoffmann, B. C. Voigt, and J. F. Poulet (2017). "Parvalbumin-expressing GABAergic neurons in primary motor cortex signal reaching". In: *Cell reports* 20.2, pp. 308–318.

Favuzzi, E., R. Deogracias, A. Marques-Smith, P. Maeso, J. Jezequel, D. Exposito-Alonso, M. Balia, T. Kroon, A. J. Hinojosa, E. F. Maraver, et al. (2019). "Distinct molecular programs regulate synapse specificity in cortical inhibitory circuits". In: *Science* 363.6425, pp. 413–417.

Ferguson, K. A. and J. A. Cardin (2020). "Mechanisms underlying gain modulation in the cortex". In: *Nature Reviews Neuroscience* 21.2, pp. 80–92.

Feulner, B. and C. Clopath (2021). "Neural manifold under plasticity in a goal driven learning behaviour". In: *PLoS computational biology* 17.2, e1008621.

Fino, E., A. M. Packer, and R. Yuste (2013). "The logic of inhibitory connectivity in the neocortex". In: *The Neuroscientist* 19.3, pp. 228–237.

Fishell, G. and A. Kepecs (2020). "Interneuron types as attractors and controllers". In: *Annual review of neuroscience* 43, pp. 1–30.

Flames, N., R. Pla, D. M. Gelman, J. L. Rubenstein, L. Puelles, and O. Marín (2007). "Delineation of multiple subpallial progenitor domains by the combinatorial expression of transcriptional codes". In: *Journal of Neuroscience* 27.36, pp. 9682–9695.

Fogarty, M., M. Grist, D. Gelman, O. Marín, V. Pachnis, and N. Kessaris (2007). "Spatial genetic patterning of the embryonic neuroepithelium generates GABAergic interneuron diversity in the adult cortex". In: *Journal of Neuroscience* 27.41, pp. 10935–10946.

Franceschetti, S., G. Sancini, F. Panzica, C. Radici, and G. Avanzini (1998). "Postnatal differentiation of firing properties and morphological characteristics in layer V pyramidal neurons of the sensorimotor cortex". In: *Neuroscience* 83.4, pp. 1013–1024.

Francioni, V., Z. Padamsey, and N. L. Rochefort (2019). "High and asymmetric somato-dendritic coupling of V1 layer 5 neurons independent of visual stimulation and locomotion". In: *Elife* 8, e49145.

Frazer, S., J. Prados, M. Niquille, C. Cadilhac, F. Markopoulos, L. Gomez, U. Tomasello, L. Telley, A. Holtmaat, D. Jabaudon, et al. (2017). "Transcriptomic and anatomic parcellation of 5-HT3AR expressing cortical interneuron subtypes revealed by single-cell RNA sequencing". In: *Nature communications* 8.1, p. 14219.

Fu, Y., J. M. Tucciarone, J. S. Espinosa, N. Sheng, D. P. Darcy, R. A. Nicoll, Z. J. Huang, and M. P. Stryker (2014). "A cortical circuit for gain control by behavioral state". In: *Cell* 156.6, pp. 1139–1152.

Fuzik, J., A. Zeisel, Z. Máté, D. Calvigioni, Y. Yanagawa, G. Szabó, S. Linnarsson, and T. Harkany (2016). "Integration of electrophysiological recordings with single-cell RNA-seq data identifies neuronal subtypes". In: *Nature biotechnology* 34.2, pp. 175–183.

Gallego, J. A., M. G. Perich, L. E. Miller, and S. A. Solla (2017). "Neural manifolds for the control of movement". In: *Neuron* 94.5, pp. 978–984.

Galvin, V. C., A. F. Arnsten, and M. Wang (2018). "Evolution in neuromodulation—the differential roles of acetylcholine in higher order association vs. primary visual cortices". In: *Frontiers in neural circuits* 12, p. 67.

García, N. V. D. M., R. Priya, S. N. Tuncdemir, G. Fishell, and T. Karayannis (2015). "Sensory inputs control the integration of neurogliaform interneurons into cortical circuits". In: *Nature neuroscience* 18.3, pp. 393–401.

Gasselin, C., B. Hohl, A. Vernet, S. Crochet, and C. C. Petersen (2021). "Cell-type-specific nicotinic input disinhibits mouse barrel cortex during active sensing". In: *Neuron* 109.5, pp. 778–787.

Geadah, V., S. Horoi, G. Kerg, G. Wolf, and G. Lajoie (2022). "Goal-driven optimization of single-neuron properties in artificial networks reveals regularization role of neural diversity and adaptation". In: *bioRxiv*, pp. 2022–04.

Gentet, L. J., Y. Kremer, H. Taniguchi, Z. J. Huang, J. F. Staiger, and C. C. Petersen (2012). "Unique functional properties of somatostatin-expressing GABAergic neurons in mouse barrel cortex". In: *Nature neuroscience* 15.4, pp. 607–612.

Gerstner, W., W. M. Kistler, R. Naud, and L. Paninski (2014). *Neuronal dynamics: From single neurons to networks and models of cognition*. Cambridge University Press.

Gidon, A., T. A. Zolnik, P. Fidzinski, F. Bolduan, A. Papoutsi, P. Poirazi, M. Holtkamp, I. Vida, and M. E. Larkum (2020). "Dendritic action potentials and computation in human layer 2/3 cortical neurons". In: *Science* 367.6473, pp. 83–87.

Goaillard, J.-M. and E. Marder (2021). "Ion channel degeneracy, variability, and covariation in neuron and circuit resilience". In: *Annual review of neuroscience* 44, pp. 335–357.

Goldberg, J. H., C. O. Lacefield, and R. Yuste (2004). "Global dendritic calcium spikes in mouse layer 5 low threshold spiking interneurones: implications for control of pyramidal cell bursting". In: *The Journal of physiology* 558.2, pp. 465–478.

Goodwin, S., J. D. McPherson, and W. R. McCombie (2016). "Coming of age: ten years of next-generation sequencing technologies". In: *Nature Reviews Genetics* 17.6, pp. 333–351.

Gould, S. J. and R. C. Lewontin (1979). "5 The Spandrels of San Marco and the Panglossian Paradigm: A Critique of the Adaptationist Programme". In: *Conceptual Issues in Evolutionary Biology* 205, p. 79.

Gould, S. J. and E. S. Vrba (1982). "Exaptation—a missing term in the science of form". In: *Paleobiology* 8.1, pp. 4–15.

Gouwens, N. W., S. A. Sorensen, F. Baftizadeh, A. Budzillo, B. R. Lee, T. Jarsky, L. Alfiler, K. Baker, E. Barkan, K. Berry, et al. (2020). "Integrated morphoelectric and transcriptomic classification of cortical GABAergic cells". In: *Cell* 183.4, pp. 935–953.

Greig, L. C., M. B. Woodworth, M. J. Galazo, H. Padmanabhan, and J. D. Macklis (2013). "Molecular logic of neocortical projection neuron specification, development and diversity". In: *Nature Reviews Neuroscience* 14.11, pp. 755–769.

Grent, T., J. Gross, J. Goense, M. Wibral, R. Gajwani, A. I. Gumley, S. M. Lawrie, M. Schwannauer, F. Schultze-Lutter, T. N. Schröder, et al. (2018). "Resting-state gamma-band power alterations in schizophrenia reveal E/I-balance abnormalities across illness-stages". In: *Elife* 7, e37799.

Guillem, K., B. Bloem, R. B. Poorthuis, M. Loos, A. B. Smit, U. Maskos, S. Spijker, and H. D. Mansvelder (2011). "Nicotinic acetylcholine receptor $\beta 2$ subunits in the medial prefrontal cortex control attention". In: *Science* 333.6044, pp. 888–891.

Gulledge, A. T., S. B. Park, Y. Kawaguchi, and G. J. Stuart (2007). "Heterogeneity of phasic cholinergic signaling in neocortical neurons". In: *Journal of neurophysiology* 97.3, pp. 2215–2229.

Güntürkün, O. and T. Bugnyar (2016). "Cognition without cortex". In: *Trends in cognitive sciences* 20.4, pp. 291–303.

Hafemeister, C. and R. Satija (2019). "Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression". In: *Genome biology* 20.1, p. 296.

Haider, B., A. Duque, A. R. Hasenstaub, and D. A. McCormick (2006). "Neocortical network activity in vivo is generated through a dynamic balance of excitation and inhibition". In: *Journal of Neuroscience* 26.17, pp. 4535–4545.

Hain, D., T. Gallego-Flores, M. Klinkmann, A. Macias, E. Ciirdaeva, A. Arends, C. Thum, G. Tushev, F. Kretschmer, M. A. Tosches, et al. (2022). "Molecular diversity and evolution of neuron types in the amniote brain". In: *Science* 377.6610, eabp8202.

Halder, G., P. Callaerts, and W. J. Gehring (1995). "Induction of ectopic eyes by targeted expression of the eyeless gene in Drosophila". In: *Science* 267.5205, pp. 1788–1792.

Hansen, D. V., J. H. Lui, P. Flandin, K. Yoshikawa, J. L. Rubenstein, A. Alvarez-Buylla, and A. R. Kriegstein (2013). "Non-epithelial stem cells and cortical interneuron production in the human ganglionic eminences". In: *Nature neuroscience* 16.11, pp. 1576–1587.

Hao, Y., S. Hao, E. Andersen-Nissen, W. M. Mauck III, S. Zheng, A. Butler, M. J. Lee, A. J. Wilk, C. Darby, M. Zager, et al. (2021). "Integrated analysis of multimodal single-cell data". In: *Cell* 184.13, pp. 3573–3587.

Haque, A., J. Engel, S. A. Teichmann, and T. Lönnberg (2017). "A practical guide to single-cell RNA-sequencing for biomedical research and clinical applications". In: *Genome medicine* 9.1, pp. 1–12.

Harris, C. R., K. J. Millman, S. J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N. J. Smith, et al. (2020). "Array programming with NumPy". In: *Nature* 585.7825, pp. 357–362.

Harris, K. D., H. Hochgerner, N. G. Skene, L. Magno, L. Katona, C. Bengtsson Gonzales, P. Somogyi, N. Kessaris, S. Linnarsson, and J. Hjerling-Leffler (2018). "Classes and continua of hippocampal CA1 inhibitory neurons revealed by single-cell transcriptomics". In: *PLoS biology* 16.6, e2006387.

Harris, K. D. and T. D. Mrsic-Flogel (2013). "Cortical connectivity and sensory coding". In: *Nature* 503.7474, pp. 51–58.

Hartung, J., A. Schroeder, R. A. Péréz Vázquez, R. B. Poorthuis, and J. J. Letzkus (2023). "Layer 1 NDNF Interneurons are Specialized Top-Down Master Regulators of Cortical Circuits". In: *bioRxiv*, pp. 2023–10.

Hasselmo, M. E. (1995). "Neuromodulation and cortical function: modeling the physiological basis of behavior". In: *Behavioural brain research* 67.1, pp. 1–27.

Hazelden, J., Y. H. Liu, E. Shlizerman, and E. Shea-Brown (2023). "Evolutionary algorithms as an alternative to backpropagation for supervised training of Biophysical Neural Networks and Neural ODEs". In: *arXiv preprint arXiv:2311.10869*.

He, M., J. Tucciarone, S. Lee, M. J. Nigro, Y. Kim, J. M. Levine, S. M. Kelly, I. Krugikov, P. Wu, Y. Chen, et al. (2016). "Strategies and tools for combinatorial targeting of GABAergic neurons in mouse cerebral cortex". In: *Neuron* 91.6, pp. 1228–1243.

Hennequin, G., E. J. Agnes, and T. P. Vogels (2017). "Inhibitory plasticity: balance, control, and codependence". In: *Annual review of neuroscience* 40, pp. 557–579.

Hennequin, G., T. P. Vogels, and W. Gerstner (2014). "Optimal control of transient dynamics in balanced networks supports generation of complex movements". In: *Neuron* 82.6, pp. 1394–1406.

Hertäg, L. and C. Clopath (2022). "Prediction-error neurons in circuits with multiple neuron types: Formation, refinement, and functional implications". In: *Proceedings of the National Academy of Sciences* 119.13, e2115699119.

Hertäg, L. and H. Sprekeler (2019). "Amplifying the redistribution of somato-dendritic inhibition by the interplay of three interneuron types". In: *PLoS computational biology* 15.5, e1006999.

Hertäg, L. and H. Sprekeler (2020). "Learning prediction error neurons in a canonical interneuron circuit". In: *bioRxiv*.

Hobert, O. (2008). "Regulatory logic of neuronal diversity: terminal selector genes and selector motifs". In: *Proceedings of the National Academy of Sciences* 105.51, pp. 20067–20071.

Hochberg, G. K. and J. W. Thornton (2017). "Reconstructing ancient proteins to understand the causes of structure and function". In: *Annual review of biophysics* 46, pp. 247–269.

Hochreiter, S. and J. Schmidhuber (1997). "Long short-term memory". In: *Neural computation* 9.8, pp. 1735–1780.

Hodge, R. D., T. E. Bakken, J. A. Miller, K. A. Smith, E. R. Barkan, L. T. Graybuck, J. L. Close, B. Long, N. Johansen, O. Penn, et al. (2019). "Conserved cell types with divergent features in human versus mouse cortex". In: *Nature* 573.7772, pp. 61–68.

Hu, H., J. Gan, and P. Jonas (2014). "Fast-spiking, parvalbumin+ GABAergic interneurons: From cellular design to microcircuit function". In: *Science* 345.6196, p. 1255263.

Hu, J. S., D. Vogt, S. Lindtner, M. Sandberg, S. N. Silberberg, and J. L. Rubenstein (2017). "Coup-TF1 and Coup-TF2 control subtype and laminar identity of MGE-derived neocortical interneurons". In: *Development* 144.15, pp. 2837–2851.

Hunter, J. D. (2007). "Matplotlib: A 2D graphics environment". In: *Computing in Science & Engineering* 9.3, pp. 90–95.

Hwang, B., J. H. Lee, and D. Bang (2018). "Single-cell RNA sequencing technologies and bioinformatics pipelines". In: *Experimental & molecular medicine* 50.8, pp. 1–14.

Iascone, D. M., Y. Li, U. Sümbül, M. Doron, H. Chen, V. Andreu, F. Goudy, H. Blockus, L. F. Abbott, I. Segev, et al. (2020). "Whole-neuron synaptic mapping reveals spatially precise excitatory/inhibitory balance limiting dendritic and somatic spiking". In: *Neuron*.

Inan, M. and S. A. Anderson (2014). "The chandelier cell, form and function". In: *Current opinion in neurobiology* 26, pp. 142–148.

Inan, M., J. Welagen, and S. A. Anderson (2012). "Spatial and temporal bias in the mitotic origins of somatostatin-and parvalbumin-expressing interneuron subgroups and the chandelier subtype in the medial ganglionic eminence". In: *Cerebral cortex* 22.4, pp. 820–827.

Isaacson, J. S. and M. Scanziani (2011). "How inhibition shapes cortical activity". In: *Neuron* 72.2, pp. 231–243.

Jaffe, P. I. and M. S. Brainard (2020). "Acetylcholine acts on songbird premotor circuitry to invigorate vocal output". In: *Elife* 9, e53288.

Jarvis, E. D., O. Güntürkün, L. Bruce, A. Csillag, H. Karten, W. Kuenzel, L. Medina, G. Paxinos, D. J. Perkel, T. Shimizu, et al. (2005). "Avian brains and a new understanding of vertebrate brain evolution". In: *Nature Reviews Neuroscience* 6.2, pp. 151–159.

Jézéquel, J., G. Condomitti, T. Kroon, F. Hamid, S. Sanalidou, T. Garces, P. Maeso, M. Balia, and B. Rico (2023). "Cadherins orchestrate specific patterns of perisomatic inhibition onto distinct pyramidal cell populations". In: *bioRxiv*, pp. 2023–09.

Jiang, X., S. Shen, C. R. Cadwell, P. Berens, F. Sinz, A. S. Ecker, S. Patel, and A. S. Tolias (2015). "Principles of connectivity among morphologically defined cell types in adult neocortex". In: *Science* 350.6264.

Johenning, F. W., P. S. Beed, T. Trimbuch, M. H. Bendels, J. Winterer, and D. Schmitz (2009). "Dendritic compartment and neuronal output mode determine pathway-specific long-term potentiation in the piriform cortex". In: *Journal of Neuroscience* 29.43, pp. 13649–13661.

Jones, D. T., W. R. Taylor, and J. M. Thornton (1992). "The rapid generation of mutation data matrices from protein sequences". In: *Bioinformatics* 8.3, pp. 275–282.

JoramKeijser (Mar. 2023a). *JoramKeijser/interneuron_evolution: Code after revision*. Version v1.0.0.

JoramKeijser (Mar. 2023b). *JoramKeijser/interneuron_evolution: Code after revision*. Version v1.0.0.

Jourjine, N. and H. E. Hoekstra (2021). "Expanding evolutionary neuroscience: insights from comparing variation in behavior". In: *Neuron* 109.7, pp. 1084–1099.

Ju, N., Y. Li, F. Liu, H. Jiang, S. L. Macknik, S. Martinez-Conde, and S. Tang (2020). "Spatiotemporal functional organization of excitatory synaptic inputs onto macaque V1 neurons". In: *Nature communications* 11.1, p. 697.

Karimi, A., J. Odenthal, F. Drawitsch, K. M. Boergens, and M. Helmstaedter (2020). "Cell-type specific innervation of cortical pyramidal cells at their apical dendrites". In: *Elife* 9, e46876.

Karnani, M. M., J. Jackson, I. Ayzenshtat, A. H. Sichani, K. Manoocheri, S. Kim, and R. Yuste (2016a). "Opening holes in the blanket of inhibition: localized lateral disinhibition by VIP interneurons". In: *Journal of neuroscience* 36.12, pp. 3471–3480.

Karnani, M. M., J. Jackson, I. Ayzenshtat, J. Tucciarone, K. Manoocheri, W. G. Snider, and R. Yuste (2016b). "Cooperative subnetworks of molecularly similar interneurons in mouse neocortex". In: *Neuron* 90.1, pp. 86–100.

Kawaguchi, Y. (1997). "Selective cholinergic modulation of cortical GABAergic cell subtypes". In: *Journal of neurophysiology* 78.3, pp. 1743–1747.

Kebschull, J. M., E. B. Richman, N. Ringach, D. Friedmann, E. Albarran, S. S. Kolluru, R. C. Jones, W. E. Allen, Y. Wang, S. W. Cho, et al. (2020). "Cerebellar nuclei evolved by repeatedly duplicating a conserved cell-type set". In: *Science* 370.6523, eabd5059.

Keijser, J. and H. Sprekeler (2023). "Cortical interneurons: fit for function and fit to function? Evidence from development and evolution". In: *Frontiers in Neural Circuits*.

Keijser, J. and H. Sprekeler (2022). "Optimizing interneuron circuits for compartment-specific feedback inhibition". In: *PLoS Computational Biology* 18.4, e1009933.

Keller, G. B. and T. D. Mrsic-Flogel (2018). "Predictive processing: a canonical cortical computation". In: *Neuron* 100.2, pp. 424–435.

Kepecs, A. and G. Fishell (2014). "Interneuron cell types are fit to function". In: *Nature* 505.7483, pp. 318–326.

Khoury, C. F., N. G. Fala, and C. A. Runyan (2022). "Region-specific modulation of somatostatin activity during arousal". In: *bioRxiv*, pp. 2022–01.

Kim, E. J., Z. Zhang, L. Huang, T. Ito-Cole, M. W. Jacobs, A. L. Juavinett, G. Senturk, M. Hu, M. Ku, J. R. Ecker, et al. (2020). "Extraction of distinct neuronal cell types from within a genetically continuous population". In: *Neuron* 107.2, pp. 274–282.

Kim, R., Y. Li, and T. J. Sejnowski (2019). "Simple framework for constructing functional spiking recurrent neural networks". In: *Proceedings of the national academy of sciences* 116.45, pp. 22811–22820.

Kim, Y., G. R. Yang, K. Pradhan, K. U. Venkataraju, M. Bota, L. C. G. Del Molino, G. Fitzgerald, K. Ram, M. He, J. M. Levine, et al. (2017). "Brain-wide maps reveal stereotyped cell-type-based cortical architecture and subcortical sexual dimorphism". In: *Cell* 171.2, pp. 456–469.

Kingma, D. P. and J. Ba (2014). "Adam: A method for stochastic optimization". In: *arXiv preprint arXiv:1412.6980*.

Kiselev, V. Y., T. S. Andrews, and M. Hemberg (2019). "Challenges in unsupervised clustering of single-cell RNA-seq data". In: *Nature Reviews Genetics* 20.5, pp. 273–282.

Klausberger, T., P. J. Magill, L. F. Márton, J. D. B. Roberts, P. M. Cobden, G. Buzsáki, and P. Somogyi (2003). "Brain-state-and cell-type-specific firing of hippocampal interneurons in vivo". In: *Nature* 421.6925, pp. 844–848.

Klos, C. and R.-M. Memmesheimer (2023). "Smooth Exact Gradient Descent Learning in Spiking Neural Networks". In: *arXiv preprint arXiv:2309.14523*.

Kolde, R. et al. (2012). "Pheatmap: pretty heatmaps". In: *R package version* 1.2, p. 726.

Kolodziejczyk, A. A., J. K. Kim, V. Svensson, J. C. Marioni, and S. A. Teichmann (2015). "The technology and biology of single-cell RNA sequencing". In: *Molecular cell* 58.4, pp. 610–620.

Körding, K. P. and P. König (2000). "Learning with two sites of synaptic integration". In: *Network: Computation in neural systems* 11.1, pp. 25–39.

Kornfeld, J., S. E. Benezra, R. T. Narayanan, F. Svara, R. Egger, M. Oberlaender, W. Denk, and M. A. Long (2017). "EM connectomics reveals axonal target variation in a sequence-generating network". In: *Elife* 6, e24364.

Köster, J. and S. Rahmann (2012). "Snakemake—a scalable bioinformatics workflow engine". In: *Bioinformatics* 28.19, pp. 2520–2522.

Kozlov, A. M., D. Darriba, T. Flouri, B. Morel, and A. Stamatakis (2019). "RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference". In: *Bioinformatics* 35.21, pp. 4453–4455.

Kriegeskorte, N. and J. Diedrichsen (2019). "Peeling the onion of brain representations". In: *Annual review of neuroscience* 42.1, pp. 407–432.

Kriegeskorte, N., M. Mur, and P. A. Bandettini (2008). "Representational similarity analysis-connecting the branches of systems neuroscience". In: *Frontiers in systems neuroscience*, p. 4.

Krienen, F. M., M. Goldman, Q. Zhang, R. C. Del Rosario, M. Florio, R. Machold, A. Saunders, K. Levandowski, H. Zaniewski, B. Schuman, et al. (2020). "Innovations present in the primate interneuron repertoire". In: *Nature*, pp. 1–8.

Krishnaswami, S. R., R. V. Grindberg, M. Novotny, P. Venepally, B. Lacar, K. Bhutani, S. B. Linker, S. Pham, J. A. Erwin, J. A. Miller, et al. (2016). "Using single nuclei for RNA-seq to capture the transcriptome of postmortem neurons". In: *Nature protocols* 11.3, pp. 499–524.

Kumar, A., O. Schiff, E. Barkai, B. W. Mel, A. Poleg-Polsky, and J. Schiller (2018). "NMDA spikes mediate amplification of inputs in the rat piriform cortex". In: *Elife* 7, e38446.

Lacar, B., S. B. Linker, B. N. Jaeger, S. R. Krishnaswami, J. J. Barron, M. J. Kelder, S. L. Parylak, A. C. Paquola, P. Venepally, M. Novotny, et al. (2016). "Nuclear RNA-seq of single neurons reveals molecular signatures of activation". In: *Nature communications* 7.1, p. 11022.

Lande, R. and S. J. Arnold (1983). "The measurement of selection on correlated characters". In: *Evolution*, pp. 1210–1226.

Larkum, M. (2013). "A cellular mechanism for cortical associations: an organizing principle for the cerebral cortex". In: *Trends in neurosciences* 36.3, pp. 141–151.

Larkum, M. E., S. Watanabe, N. Lasser-Ross, P. Rhodes, and W. N. Ross (2008). "Dendritic properties of turtle pyramidal neurons". In: *Journal of neurophysiology* 99.2, pp. 683–694.

Larkum, M. E., J. J. Zhu, and B. Sakmann (1999). "A new cellular mechanism for coupling inputs arriving at different cortical layers". In: *Nature* 398.6725, pp. 338–341.

Larkum, M. E., J. J. Zhu, and B. Sakmann (2001). "Dendritic mechanisms underlying the coupling of the dendritic with the axonal action potential initiation zone of adult rat layer 5 pyramidal neurons". In: *The Journal of physiology* 533.2, pp. 447–466.

Laurent, G. (2020). "On the value of model diversity in neuroscience". In: *Nature Reviews Neuroscience* 21.8, pp. 395–396.

Lause, J., P. Berens, and D. Kobak (2021). "Analytic Pearson residuals for normalization of single-cell RNA-seq UMI data". In: *Genome biology* 22.1, pp. 1–20.

LeCun, Y., Y. Bengio, and G. Hinton (2015). "Deep learning". In: *nature* 521.7553, pp. 436–444.

Ledderose, J. M., T. A. Zolnik, M. Toumazou, T. Trimbuch, C. Rosenmund, B. J. Eickholt, D. Jaeger, M. E. Larkum, and R. N. Sachdev (2022). "Input to cortical layer 1 of somatosensory cortex". In: *bioRxiv*, pp. 2021–11.

Ledderose, J. M., T. A. Zolnik, M. Toumazou, T. Trimbuch, C. Rosenmund, B. J. Eickholt, D. Jaeger, M. E. Larkum, and R. N. Sachdev (2023). "Layer 1 of somatosensory cortex: an important site for input to a tiny cortical compartment". In: *Cerebral Cortex*, bhad371.

Lee, S., I. Kruglikov, Z. J. Huang, G. Fishell, and B. Rudy (2013). "A disinhibitory circuit mediates motor integration in the somatosensory cortex". In: *Nature neuroscience* 16.11, pp. 1662–1670.

Lein, E., L. E. Borm, and S. Linnarsson (2017). "The promise of spatial transcriptomics for neuroscience in the era of molecular cell typing". In: *Science* 358.6359, pp. 64–69.

Letzkus, J. J., B. M. Kampa, and G. J. Stuart (2006). "Learning rules for spike timing-dependent plasticity depend on dendritic synapse location". In: *Journal of Neuroscience* 26.41, pp. 10420–10429.

Letzkus, J. J., S. B. Wolff, and A. Lüthi (2015). "Disinhibition, a circuit mechanism for associative learning and memory". In: *Neuron* 88.2, pp. 264–276.

Li, H. and N. Homer (2010). "A survey of sequence alignment algorithms for next-generation sequencing". In: *Briefings in bioinformatics* 11.5, pp. 473–483.

Li, X., B. Yu, Q. Sun, Y. Zhang, M. Ren, X. Zhang, A. Li, J. Yuan, L. Madisen, Q. Luo, et al. (2018). "Generation of a whole-brain atlas for the cholinergic system and mesoscopic projectome analysis of basal forebrain cholinergic neurons". In: *Proceedings of the National Academy of Sciences* 115.2, pp. 415–420.

Liguz-Lecznar, M., J. Urban-Ciecko, and M. Kossut (2016). "Somatostatin and somatostatin-containing neurons in shaping neuronal activity and plasticity". In: *Frontiers in neural circuits* 10, p. 48.

Lillicrap, T. P. and A. Santoro (2019). "Backpropagation through time and the brain". In: *Current opinion in neurobiology* 55, pp. 82–89.

Lim, L., D. Mi, A. Llorca, and O. Marín (2018a). "Development and functional diversification of cortical interneurons". In: *Neuron* 100.2, pp. 294–313.

Lim, L., J. M. Pakan, M. M. Selten, A. Marques-Smith, A. Llorca, S. E. Bae, N. L. Rochefort, and O. Marín (2018b). "Optimization of interneuron function by direct coupling of cell migration and axonal targeting". In: *Nature neuroscience* 21.7, pp. 920–931.

Lipovsek, M., C. Bardy, C. R. Cadwell, K. Hadley, D. Kobak, and S. J. Tripathy (2021). "Patch-seq: Past, present, and future". In: *Journal of Neuroscience* 41.5, pp. 937–946.

Litwin-Kumar, A., R. Rosenbaum, and B. Doiron (2016). "Inhibitory stabilization and visual coding in cortical circuits with multiple interneuron subtypes". In: *Journal of neurophysiology* 115.3, pp. 1399–1409.

Lodato, S., B. J. Molyneaux, E. Zuccaro, L. A. Goff, H.-H. Chen, W. Yuan, A. Meleski, E. Takahashi, S. Mahony, J. L. Rinn, et al. (2014). "Gene co-regulation by Fezf2 selects neurotransmitter identity and connectivity of corticospinal neurons". In: *Nature neuroscience* 17.8, pp. 1046–1054.

Lodato, S., C. Rouaux, K. B. Quast, C. Jantrachotechatchawan, M. Studer, T. K. Hensch, and P. Arlotta (2011). "Excitatory projection neuron subtypes control the distribution of local inhibitory interneurons in the cerebral cortex". In: *Neuron* 69.4, pp. 763–779.

London, M. and M. Häusser (2005). "Dendritic computation". In: *Annu. Rev. Neurosci.* 28, pp. 503–532.

Lovett-Barron, M., A. S. Andalman, W. E. Allen, S. Vesuna, I. Kauvar, V. M. Burns, and K. Deisseroth (2017). "Ancestral circuits for the coordinated modulation of brain state". In: *Cell* 171.6, pp. 1411–1423.

Lovett-Barron, M., G. F. Turi, P. Kaifosh, P. H. Lee, F. Bolze, X.-H. Sun, J.-F. Nicoud, B. V. Zemelman, S. M. Sternson, and A. Losonczy (2012). "Regulation of neuronal input transformations by tunable dendritic inhibition". In: *Nature neuroscience* 15.3, pp. 423–430.

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

Ma, T., C. Wang, L. Wang, X. Zhou, M. Tian, Q. Zhang, Y. Zhang, J. Li, Z. Liu, Y. Cai, et al. (2013). "Subcortical origins of human and monkey neocortical interneurons". In: *Nature neuroscience* 16.11, pp. 1588–1597.

Mackwood, O., L. B. Naumann, and H. Sprekeler (2021). "Learning excitatory-inhibitory neuronal assemblies in recurrent networks". In: *Elife* 10, e59715.

Macosko, E. Z., A. Basu, R. Satija, J. Nemesh, K. Shekhar, M. Goldman, I. Tirosh, A. R. Bialas, N. Kamitaki, E. M. Martersteck, et al. (2015). "Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets". In: *Cell* 161.5, pp. 1202–1214.

Malina, K. C.-K., E. Tsivourakis, D. Kushinsky, D. Apelblat, S. Shtiglitz, E. Zohar, M. Sokoletsky, G.-i. Tasaka, A. Mizrahi, I. Lampl, et al. (2021). "NDNF interneurons in layer 1 gain-modulate whole cortical columns according to an animal's behavioral state". In: *Neuron* 109.13, pp. 2150–2164.

Mante, V., D. Sussillo, K. V. Shenoy, and W. T. Newsome (2013). "Context-dependent computation by recurrent dynamics in prefrontal cortex". In: *nature* 503.7474, pp. 78–84.

Marder, E., T. O'Leary, and S. Shruti (2014). "Neuromodulation of circuits with variable parameters: single neurons and small circuits reveal principles of state-dependent and robust neuromodulation". In: *Annual review of neuroscience* 37, pp. 329–346.

Marín, O. and J. L. Rubenstein (2001). "A long, remarkable journey: tangential migration in the telencephalon". In: *Nature Reviews Neuroscience* 2.11, pp. 780–790.

Markram, H., M. Toledo-Rodriguez, Y. Wang, A. Gupta, G. Silberberg, and C. Wu (2004). "Interneurons of the neocortical inhibitory system". In: *Nature reviews neuroscience* 5.10, pp. 793–807.

Markram, H., Y. Wang, and M. Tsodyks (1998). "Differential signaling via the same axon of neocortical pyramidal neurons". In: *Proceedings of the National Academy of Sciences* 95.9, pp. 5323–5328.

Marques-Smith, A., D. Lyngholm, A.-K. Kaufmann, J. A. Stacey, A. Hoerder-Suabedissen, E. B. Becker, M. C. Wilson, Z. Molnár, and S. J. Butt (2016). "A transient translaminar GABAergic interneuron circuit connects thalamocortical recipient layers in neonatal somatosensory cortex". In: *Neuron* 89.3, pp. 536–549.

Mason, P. H., B. Winter, A. Grignolio, et al. (2015). "Hidden in plain view: degeneracy in complex systems". In: *Biosystems* 128, pp. 1–8.

Masse, N. Y., G. R. Yang, H. F. Song, X.-J. Wang, and D. J. Freedman (2019). "Circuit mechanisms for the maintenance and manipulation of information in working memory". In: *Nature neuroscience* 22.7, pp. 1159–1167.

Mastrogiuseppe, F. and S. Ostojic (2018). "Linking connectivity, dynamics, and computations in low-rank recurrent neural networks". In: *Neuron* 99.3, pp. 609–623.

Mathis, A., P. Mamidanna, K. M. Cury, T. Abe, V. N. Murthy, M. W. Mathis, and M. Bethge (2018). "DeepLabCut: markerless pose estimation of user-defined body parts with deep learning". In: *Nature neuroscience* 21.9, pp. 1281–1289.

Mayer, C., C. Hafemeister, R. C. Bandler, R. Machold, R. B. Brito, X. Jaglin, K. Allaway, A. Butler, G. Fishell, and R. Satija (2018). "Developmental diversification of cortical inhibitory interneurons". In: *Nature* 555.7697, pp. 457–462.

McCormick, D. A., D. B. Nestvogel, and B. J. He (2020). "Neuromodulation of brain state and behavior". In: *Annual review of neuroscience* 43, pp. 391–415.

McDonald, J. H. and M. Kreitman (1991). "Adaptive protein evolution at the Adh locus in Drosophila". In: *Nature* 351.6328, pp. 652–654.

McInnes, L., J. Healy, and J. Melville (2018). "Umap: Uniform manifold approximation and projection for dimension reduction". In: *arXiv preprint arXiv:1802.03426*.

McKenzie, M. G., L. V. Cobbs, P. D. Dummer, T. J. Petros, M. M. Halford, S. A. Stacker, Y. Zou, G. J. Fishell, and E. Au (2019). "Non-canonical Wnt signaling through Ryk regulates the generation of somatostatin-and parvalbumin-expressing cortical interneurons". In: *Neuron* 103.5, pp. 853–864.

Melzer, S. and H. Monyer (2020). "Diversity and function of corticopetal and corticofugal GABAergic projection neurons". In: *Nature Reviews Neuroscience* 21.9, pp. 499–515.

Metherate, R., C. L. Cox, and J. H. Ashe (1992). "Cellular bases of neocortical activation: modulation of neural oscillations by the nucleus basalis and endogenous acetylcholine". In: *Journal of Neuroscience* 12.12, pp. 4701–4711.

Metzker, M. L. (2010). "Sequencing technologies—the next generation". In: *Nature reviews genetics* 11.1, pp. 31–46.

Mi, D., Z. Li, L. Lim, M. Li, M. Moissidis, Y. Yang, T. Gao, T. X. Hu, T. Pratt, D. J. Price, et al. (2018). "Early emergence of cortical interneuron diversity in the mouse embryo". In: *Science* 360.6384, pp. 81–85.

Mineault, P. and K. Nozawa (2021). *patrickmineault/codebook: 1.0.0*.

Miyoshi, G., J. Hjerling-Leffler, T. Karayannis, V. H. Sousa, S. J. Butt, J. Battiste, J. E. Johnson, R. P. Machold, and G. Fishell (2010). "Genetic fate mapping reveals that the caudal ganglionic eminence produces a large and diverse population of superficial cortical interneurons". In: *Journal of Neuroscience* 30.5, pp. 1582–1594.

Molano-Mazón, M., Y. Shao, D. Duque, G. R. Yang, S. Ostojic, and J. de la Rocha (2023). "Recurrent networks endowed with structural priors explain suboptimal animal behavior". In: *Current Biology* 33.4, pp. 622–638.

Moran, R. J., P. Campo, M. Symmonds, K. E. Stephan, R. J. Dolan, and K. J. Friston (2013). "Free energy, precision and learning: the role of cholinergic neuromodulation". In: *Journal of Neuroscience* 33.19, pp. 8227–8236.

Morel, B., A. M. Kozlov, A. Stamatakis, and G. J. Szöllősi (2020). "GeneRax: a tool for species-tree-aware maximum likelihood-based gene family tree inference under gene duplication, transfer, and loss". In: *Molecular biology and evolution* 37.9, pp. 2763–2774.

Muñoz, W., R. Tremblay, D. Levenstein, and B. Rudy (2017). "Layer-specific modulation of neocortical dendritic inhibition during active wakefulness". In: *Science* 355.6328, pp. 954–959.

Murayama, M., E. Pérez-Garci, T. Nevian, T. Bock, W. Senn, and M. E. Larkum (2009). "Dendritic encoding of sensory stimuli controlled by deep cortical interneurons". In: *Nature* 457.7233, pp. 1137–1141.

Murphy, B. K. and K. D. Miller (2009). "Balanced amplification: a new mechanism of selective amplification of neural activity patterns". In: *Neuron* 61.4, pp. 635–648.

Naka, A., J. Veit, B. Shababo, R. K. Chance, D. Risso, D. Stafford, B. Snyder, A. Egladyous, D. Chu, S. Sridharan, et al. (2019). "Complementary networks of cortical somatostatin interneurons enforce layer specific control". In: *Elife* 8, e43696.

Nashef, A., O. Cohen, S. I. Perlmutter, and Y. Prut (2022). "A cerebellar origin of feedforward inhibition to the motor cortex in non-human primates". In: *Cell Reports* 39.6.

Naskar, S., J. Qi, F. Pereira, C. R. Gerfen, and S. Lee (2021). "Cell-type-specific recruitment of GABAergic interneurons in the primary somatosensory cortex by long-range inputs". In: *Cell reports* 34.8.

Naud, R., B. Bathellier, and W. Gerstner (2013). "Spike timing prediction with active dendrites". In: *arXiv preprint arXiv:1311.3586*.

Naud, R., B. Bathellier, and W. Gerstner (2014). "Spike-timing prediction in cortical neurons with active dendrites". In: *Frontiers in computational neuroscience* 8, p. 90.

Naud, R. and H. Sprekeler (2018). "Sparse bursts optimize information transmission in a multiplexed neural code". In: *Proceedings of the National Academy of Sciences* 115.27, E6329–E6338.

Naumann, L. B., J. Keijser, and H. Sprekeler (2022). "Invariant neural subspaces maintained by feedback modulation". In: *eLife* 11, e76096.

Naumann, R. K., J. M. Ondracek, S. Reiter, M. Shein-Idelson, M. A. Tosches, T. M. Yamawaki, and G. Laurent (2015). "The reptilian brain". In: *Current Biology* 25.8, R317–R321.

Neftci, E. O., H. Mostafa, and F. Zenke (2019). "Surrogate gradient learning in spiking neural networks: Bringing the power of gradient-based optimization to spiking neural networks". In: *IEEE Signal Processing Magazine* 36.6, pp. 51–63.

Nery, S., G. Fishell, and J. G. Corbin (2002). "The caudal ganglionic eminence is a source of distinct cortical and subcortical cell populations". In: *Nature neuroscience* 5.12, pp. 1279–1287.

Nigro, M. J., Y. Hashikawa-Yamasaki, and B. Rudy (2018). "Diversity and connectivity of layer 5 somatostatin-expressing interneurons in the mouse barrel cortex". In: *Journal of Neuroscience* 38.7, pp. 1622–1633.

Nomura, T., W. Yamashita, H. Gotoh, and K. Ono (2018). "Species-specific mechanisms of neuron subtype specification reveal evolutionary plasticity of amniote brain development". In: *Cell reports* 22.12, pp. 3142–3151.

O'Donnell, P. (2011). "Adolescent onset of cortical disinhibition in schizophrenia: insights from animal models". In: *Schizophrenia bulletin* 37.3, pp. 484–492.

Obermayer, J., M. B. Verhoog, A. Luchicchi, and H. D. Mansvelder (2017). "Cholinergic modulation of cortical microcircuits is layer-specific: evidence from rodent, monkey and human brain". In: *Frontiers in neural circuits* 11, p. 100.

Okun, M. and I. Lampl (2008). "Instantaneous correlation of excitation and inhibition during ongoing and sensory-evoked activities". In: *Nature neuroscience* 11.5, pp. 535–537.

Orhan, A. E. and W. J. Ma (2019). "A diverse range of factors affect the nature of neural representations underlying short-term memory". In: *Nature neuroscience* 22.2, pp. 275–283.

Orlandi, K. N., S. R. Phillips, Z. R. Sailer, J. L. Harman, and M. J. Harms (2023). "Topiary: Pruning the manual labor from ancestral sequence reconstruction". In: *Protein Science* 32.2, e4551.

Ozeki, H., I. M. Finn, E. S. Schaffer, K. D. Miller, and D. Ferster (2009). "Inhibitory stabilization of the cortical network underlies visual surround suppression". In: *Neuron* 62.4, pp. 578–592.

Packer, A. M. and R. Yuste (2011). "Dense, unspecific connectivity of neocortical parvalbumin-positive interneurons: a canonical microcircuit for inhibition?" In: *Journal of Neuroscience* 31.37, pp. 13260–13271.

Pandarinath, C., D. J. O'Shea, J. Collins, R. Jozefowicz, S. D. Stavisky, J. C. Kao, E. M. Trautmann, M. T. Kaufman, S. I. Ryu, L. R. Hochberg, et al. (2018). "Inferring single-trial neural population dynamics using sequential auto-encoders". In: *Nature methods* 15.10, pp. 805–815.

Parikh, V., R. Kozak, V. Martinez, and M. Sarter (2007). "Prefrontal acetylcholine release controls cue detection on multiple timescales". In: *Neuron* 56.1, pp. 141–154.

Pascanu, R., T. Mikolov, and Y. Bengio (2013). "On the difficulty of training recurrent neural networks". In: *International conference on machine learning*, pp. 1310–1318.

Paszke, A., S. Gross, S. Chintala, G. Chanan, E. Yang, Z. DeVito, Z. Lin, A. Desmaison, L. Antiga, and A. Lerer (2017). "Automatic differentiation in pytorch". In.

Paszke, A., S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, et al. (2019a). "Pytorch: An imperative style, high-performance deep learning library". In: *Advances in neural information processing systems* 32.

Paszke, A. et al. (2019b). "PyTorch: An Imperative Style, High-Performance Deep Learning Library". In: *Advances in Neural Information Processing Systems 32*. Ed. by H. Wallach, H. Larochelle, A. Beygelzimer, F. d' Alché-Buc, E. Fox, and R. Garnett. Curran Associates, Inc., pp. 8024–8035.

Paul, A., M. Crow, R. Raudales, M. He, J. Gillis, and Z. J. Huang (2017). "Transcriptional architecture of synaptic communication delineates GABAergic neuron identity". In: *Cell* 171.3, pp. 522–539.

Payeur, A., J.-C. Béique, and R. Naud (2019). "Classes of dendritic information processing". In: *Current opinion in neurobiology* 58, pp. 78–85.

Payeur, A., J. Guerguiev, F. Zenke, B. A. Richards, and R. Naud (2021). "Burst-dependent synaptic plasticity can coordinate learning in hierarchical circuits". In: *Nature neuroscience* 24.7, pp. 1010–1019.

Pedregosa, F. et al. (2011). "Scikit-learn: Machine Learning in Python". In: *Journal of Machine Learning Research* 12, pp. 2825–2830.

Peng, H., P. Xie, L. Liu, X. Kuang, Y. Wang, L. Qu, H. Gong, S. Jiang, A. Li, Z. Ruan, et al. (2021). "Morphological diversity of single neurons in molecularly defined cell types". In: *Nature* 598.7879, pp. 174–181.

Pereira, T. D., J. W. Shaevitz, and M. Murthy (2020). "Quantifying behavior to understand the brain". In: *Nature neuroscience* 23.12, pp. 1537–1549.

Perez-Nieves, N., V. C. Leung, P. L. Dragotti, and D. F. Goodman (2021). "Neural heterogeneity promotes robust learning". In: *Nature communications* 12.1, p. 5791.

Perich, M. G., C. Arlt, S. Soares, M. E. Young, C. P. Mosher, J. Minxha, E. Carter, U. Rutishauser, P. H. Rudebeck, C. D. Harvey, et al. (2020). "Inferring brain-wide interactions using data-constrained recurrent neural network models". In: *BioRxiv*, pp. 2020–12.

Petreanu, L., D. Huber, A. Sobczyk, and K. Svoboda (2007). "Channelrhodopsin-2–assisted circuit mapping of long-range callosal projections". In: *Nature neuroscience* 10.5, pp. 663–668.

Petreanu, L., T. Mao, S. M. Sternson, and K. Svoboda (2009). "The subcellular organization of neocortical excitatory connections". In: *Nature* 457.7233, pp. 1142–1145.

Pfeffer, C. K., M. Xue, M. He, Z. J. Huang, and M. Scanziani (2013). "Inhibition of inhibition in visual cortex: the logic of connections between molecularly distinct interneurons". In: *Nature neuroscience* 16.8, pp. 1068–1076.

Pi, H.-J., B. Hangya, D. Kvitsiani, J. I. Sanders, Z. J. Huang, and A. Kepecs (2013). "Cortical interneurons that specialize in disinhibitory control". In: *Nature* 503.7477, pp. 521–524.

Picelli, S., O. R. Faridani, Å. K. Björklund, G. Winberg, S. Sagasser, and R. Sandberg (2014). "Full-length RNA-seq from single cells using Smart-seq2". In: *Nature protocols* 9.1, pp. 171–181.

Pigliucci, M. and J. Kaplan (2000). "The fall and rise of Dr Pangloss: adaptationism and the Spandrels paper 20 years later". In: *Trends in ecology & evolution* 15.2, pp. 66–70.

Pinto, L., M. J. Goard, D. Estandian, M. Xu, A. C. Kwan, S.-H. Lee, T. C. Harrison, G. Feng, and Y. Dan (2013). "Fast modulation of visual perception by basal forebrain cholinergic neurons". In: *Nature neuroscience* 16.12, pp. 1857–1863.

Poirazi, P., T. Brannon, and B. W. Mel (2003). "Pyramidal neuron as two-layer neural network". In: *Neuron* 37.6, pp. 989–999.

Poirazi, P. and A. Papoutsi (2020). "Illuminating dendritic function with computational models". In: *Nature Reviews Neuroscience*, pp. 1–19.

Polsky, A., B. W. Mel, and J. Schiller (2004). "Computational subunits in thin dendrites of pyramidal cells". In: *Nature neuroscience* 7.6, pp. 621–627.

Poorthuis, R. B., B. Bloem, B. Schak, J. Wester, C. P. de Kock, and H. D. Mansvelder (2013). "Layer-specific modulation of the prefrontal cortex by nicotinic acetylcholine receptors". In: *Cerebral cortex* 23.1, pp. 148–161.

Poorthuis, R. B., L. Enke, and J. J. Letzkus (2014). "Cholinergic circuit modulation through differential recruitment of neocortical interneuron types during behaviour". In: *The Journal of physiology* 592.19, pp. 4155–4164.

Pouille, F. and M. Scanziani (2004). "Routing of spike series by dynamic circuits in the hippocampus". In: *Nature* 429.6993, pp. 717–723.

Prinz, A. A., D. Bucher, and E. Marder (2004). "Similar network activity from disparate circuit parameters". In: *Nature neuroscience* 7.12, pp. 1345–1352.

Puzerey, P. A., K. Maher, N. Prasad, and J. H. Goldberg (2018). "Vocal learning in songbirds requires cholinergic signaling in a motor cortex-like nucleus". In: *Journal of neurophysiology* 120.4, pp. 1796–1806.

Qian, X., K. D. Harris, T. Hauling, D. Nicoloutsopoulos, A. B. Muñoz-Manchado, N. Skene, J. Hjerling-Leffler, and M. Nilsson (2020). "Probabilistic cell typing enables fine mapping of closely related cell types in situ". In: *Nature methods* 17.1, pp. 101–106.

Raj, A., C. S. Peskin, D. Tranchina, D. Y. Vargas, and S. Tyagi (2006). "Stochastic mRNA synthesis in mammalian cells". In: *PLoS biology* 4.10, e309.

Rakic, P. (1971). "Neuron-glia relationship during granule cell migration in developing cerebellar cortex. A Golgi and electonmicroscopic study in Macacus rhesus". In: *Journal of Comparative Neurology* 141.3, pp. 283–312.

Ramón y Cajal, S. (1894). "The Croonian lecture.—La fine structure des centres nerveux". In: *Proceedings of the Royal Society of London* 55.331-335, pp. 444–468.

Rao, R. P. and D. H. Ballard (1999). "Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects". In: *Nature neuroscience* 2.1, pp. 79–87.

Reiner, A. (1993). "Neurotransmitter organization and connections of turtle cortex: implications for the evolution of mammalian isocortex". In: *Comparative Biochemistry and Physiology Part A: Physiology* 104.4, pp. 735–748.

Reiner, A., D. J. Perkel, L. L. Bruce, A. B. Butler, A. Csillag, W. Kuenzel, L. Medina, G. Paxinos, T. Shimizu, G. Striedter, et al. (2004). "Revised nomenclature for avian telencephalon and some related brainstem nuclei". In: *Journal of Comparative Neurology* 473.3, pp. 377–414.

Renart, A., J. De La Rocha, P. Bartho, L. Hollender, N. Parga, A. Reyes, and K. D. Harris (2010). "The asynchronous state in cortical circuits". In: *science* 327.5965, pp. 587–590.

Reyes, A., R. Lujan, A. Rozov, N. Burnashev, P. Somogyi, and B. Sakmann (1998). "Target-cell-specific facilitation and depression in neocortical circuits". In: *Nature neuroscience* 1.4, pp. 279–285.

Richards, B. A., T. P. Lillicrap, P. Beaudoin, Y. Bengio, R. Bogacz, A. Christensen, C. Clopath, R. P. Costa, A. de Berker, S. Ganguli, et al. (2019). "A deep learning framework for neuroscience". In: *Nature neuroscience* 22.11, pp. 1761–1770.

Rodriques, S. G., R. R. Stickels, A. Goeva, C. A. Martin, E. Murray, C. R. Vanderburg, J. Welch, L. M. Chen, F. Chen, and E. Z. Macosko (2019). "Slide-seq: A scalable technology for measuring genome-wide expression at high spatial resolution". In: *Science* 363.6434, pp. 1463–1467.

Romand, S., Y. Wang, M. Toledo-Rodriguez, and H. Markram (2011). "Morphological development of thick-tufted layer v pyramidal cells in the rat somatosensory cortex". In: *Frontiers in neuroanatomy* 5, p. 5.

Rossbroich, J., D. Trotter, J. Beninger, K. Tóth, and R. Naud (2021). "Linear-nonlinear cascades capture synaptic dynamics". In: *PLoS computational biology* 17.3, e1008013.

Royer, S., B. V. Zemelman, A. Losonczy, J. Kim, F. Chance, J. C. Magee, and G. Buzsáki (2012). "Control of timing, rate and bursts of hippocampal place cells by dendritic and somatic inhibition". In: *Nature neuroscience* 15.5, pp. 769–775.

Rubin, D. B., S. D. Van Hooser, and K. D. Miller (2015). "The stabilized supralinear network: a unifying circuit motif underlying multi-input integration in sensory cortex". In: *Neuron* 85.2, pp. 402–417.

Rudy, B., G. Fishell, S. Lee, and J. Hjerling-Leffler (2011). "Three groups of interneurons account for nearly 100% of neocortical GABAergic neurons". In: *Developmental neurobiology* 71.1, pp. 45–61.

Rumelhart, D. E., G. E. Hinton, and R. J. Williams (1986). "Learning representations by back-propagating errors". In: *nature* 323.6088, pp. 533–536.

Rupprecht, P. and R. W. Friedrich (2018). "Precise synaptic balance in the zebrafish homolog of olfactory cortex". In: *Neuron* 100.3, pp. 669–683.

Sacramento, J., R. Ponte Costa, Y. Bengio, and W. Senn (2018). "Dendritic cortical microcircuits approximate the backpropagation algorithm". In: *Advances in neural information processing systems* 31.

Sadeh, S. and C. Clopath (2021). "Inhibitory stabilization and cortical computation". In: *Nature Reviews Neuroscience* 22.1, pp. 21–37.

Sanzeni, A., B. Akitake, H. C. Goldbach, C. E. Leedy, N. Brunel, and M. H. Histed (2020). "Inhibition stabilization is a widespread property of cortical networks". In: *Elife* 9, e54875.

Saxe, A. M., J. L. McClelland, and S. Ganguli (2019). "A mathematical theory of semantic development in deep neural networks". In: *Proceedings of the National Academy of Sciences* 116.23, pp. 11537–11546.

Scala, F., D. Kobak, M. Bernabucci, Y. Bernaerts, C. R. Cadwell, J. R. Castro, L. Hartmanis, X. Jiang, S. Laturnus, E. Miranda, et al. (2020). "Phenotypic variation of transcriptomic cell types in mouse motor cortex". In: *Nature*, pp. 1–7.

Scala, F., D. Kobak, M. Bernabucci, Y. Bernaerts, C. R. Cadwell, J. R. Castro, L. Hartmanis, X. Jiang, S. Laturnus, E. Miranda, et al. (2021). "Phenotypic variation of transcriptomic cell types in mouse motor cortex". In: *Nature* 598.7879, pp. 144–150.

Scala, F., D. Kobak, S. Shan, Y. Bernaerts, S. Laturnus, C. R. Cadwell, L. Hartmanis, E. Froudarakis, J. R. Castro, Z. H. Tan, et al. (2019). "Layer 4 of mouse neocortex differs in cell types and circuit organization between sensory areas". In: *Nature communications* 10.1, p. 4174.

Schaeffer, R., M. Khona, T. Ma, C. Eyzaguirre, S. Koyejo, and I. R. Fiete (2023). "Self-Supervised Learning of Representations for Space Generates Multi-Modular Grid Cells". In: *arXiv preprint arXiv:2311.02316*.

Schmitz, M. T., K. Sandoval, C. P. Chen, M. A. Mostajo-Radji, W. W. Seeley, T. J. Nowakowski, C. J. Ye, M. F. Paredes, and A. A. Pollen (2022). "The development and evolution of inhibitory neurons in primate cerebrum". In: *Nature* 603.7903, pp. 871–877.

Schroeder, A., M. B. Pardi, J. Keijser, T. Dalmay, A. I. Groisman, E. M. Schuman, H. Sprekeler, and J. J. Letzkus (2023). "Inhibitory top-down projections from zona incerta mediate neocortical memory". In: *Neuron*.

Seabold, S. and J. Perktold (2010). "statsmodels: Econometric and statistical modeling with python". In: *9th Python in Science Conference*.

Seignette, K., N. Jamann, P. Papale, H. Terra, R. P. Porneso, L. de Kraker, C. van der Togt, M. van der Aa, P. Neering, E. Ruimschotel, et al. (2023). "Visuomotor experience induces functional and structural plasticity of chandelier cells". In: *bioRxiv*, pp. 2023–04.

Shai, A. S., C. A. Anastassiou, M. E. Larkum, and C. Koch (2015). "Physiology of layer 5 pyramidal neurons in mouse primary visual cortex: coincidence detection through bursting". In: *PLoS computational biology* 11.3, e1004090.

Shen, S., X. Jiang, F. Scala, J. Fu, P. Fahey, D. Kobak, Z. Tan, N. Zhou, J. Reimer, F. Sinz, et al. (2022). "Distinct organization of two cortico-cortical feedback pathways". In: *Nature Communications* 13.1, p. 6389.

Shi, Y., M. Wang, D. Mi, T. Lu, B. Wang, H. Dong, S. Zhong, Y. Chen, L. Sun, X. Zhou, et al. (2021). "Mouse and human share conserved transcriptional programs for interneuron development". In: *Science* 374.6573, eabj6641.

Silberberg, G. and H. Markram (2007). "Disynaptic inhibition between neocortical pyramidal cells mediated by Martinotti cells". In: *Neuron* 53.5, pp. 735–746.

Sjostrom, P. J., E. A. Rancz, A. Roth, and M. Hausser (2008). "Dendritic excitability and synaptic plasticity". In: *Physiological reviews* 88.2, pp. 769–840.

Smith, J. M. (1990). "Optimality theory in evolutionary biology". In: *Nature* 348.6296, pp. 27–33.

Sohal, V. S. and J. L. Rubenstein (2019). "Excitation-inhibition balance as a framework for investigating mechanisms in neuropsychiatric disorders". In: *Molecular psychiatry* 24.9, pp. 1248–1257.

Sohn, H., D. Narain, N. Meirhaeghe, and M. Jazayeri (2019). "Bayesian computation through cortical latent dynamics". In: *Neuron* 103.5, pp. 934–947.

Somogyi, P. (1977). "A specific 'axo-axonal'interneuron in the visual cortex of the rat". In: *Brain Res* 136.2, pp. 345–350.

Soneson, C. and M. D. Robinson (2018). "Bias, robustness and scalability in single-cell differential expression analysis". In: *Nature methods* 15.4, pp. 255–261.

Song, H. F., G. R. Yang, and X.-J. Wang (2016). "Training excitatory-inhibitory recurrent neural networks for cognitive tasks: a simple and flexible framework". In: *PLoS computational biology* 12.2, e1004792.

Song, S., K. D. Miller, and L. F. Abbott (2000). "Competitive Hebbian learning through spike-timing-dependent synaptic plasticity". In: *Nature neuroscience* 3.9, pp. 919–926.

Stachniak, T. J., R. Kastli, O. Hanley, A. Ö. Argunsah, E. G. T. van der Valk, G. Kanatouris, and T. Karayannis (2021). "Postmitotic Prox1 Expression Controls the Final Specification of Cortical VIP Interneuron Subtypes". In: *Journal of Neuroscience* 41.39, pp. 8150–8162.

Stachniak, T. J., E. L. Sylwestrak, P. Scheiffele, B. J. Hall, and A. Ghosh (2019). "Elfn1-induced constitutive activation of mGluR7 determines frequency-dependent recruitment of somatostatin interneurons". In: *Journal of Neuroscience* 39.23, pp. 4461–4474.

Stark, R., M. Grzelak, and J. Hadfield (2019). "RNA sequencing: the teenage years". In: *Nature Reviews Genetics* 20.11, pp. 631–656.

Stefanelli, T., C. Bertollini, C. Lüscher, D. Muller, and P. Mendez (2016). "Hippocampal somatostatin interneurons control the size of neuronal memory ensembles". In: *Neuron* 89.5, pp. 1074–1085.

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

Stickels, R. R., E. Murray, P. Kumar, J. Li, J. L. Marshall, D. J. Di Bella, P. Arlotta, E. Z. Macosko, and F. Chen (2021). "Highly sensitive spatial transcriptomics at near-cellular resolution with Slide-seqV2". In: *Nature biotechnology* 39.3, pp. 313–319.

Stuart, G., J. Schiller, and B. Sakmann (1997). "Action potential initiation and propagation in rat neocortical pyramidal neurons". In: *The Journal of physiology* 505.3, pp. 617–632.

Stuart, G. J. and B. Sakmann (1994). "Active propagation of somatic action potentials into neocortical pyramidal cell dendrites". In: *Nature* 367.6458, pp. 69–72.

Stuart, T., A. Butler, P. Hoffman, C. Hafemeister, E. Papalexi, W. M. Mauck, Y. Hao, M. Stoeckius, P. Smibert, and R. Satija (2019). "Comprehensive integration of single-cell data". In: *Cell* 177.7, pp. 1888–1902.

Sussel, L., O. Marin, S. Kimura, and J. Rubenstein (1999). "Loss of Nkx2. 1 homeobox gene function results in a ventral to dorsal molecular respecification within the basal telencephalon: evidence for a transformation of the pallidum into the striatum". In: *Development* 126.15, pp. 3359–3370.

Sussillo, D. (2014). "Neural circuits as computational dynamical systems". In: *Current opinion in neurobiology* 25, pp. 156–163.

Sussillo, D., M. M. Churchland, M. T. Kaufman, and K. V. Shenoy (2015). "A neural network that finds a naturalistic solution for the production of muscle activity". In: *Nature neuroscience* 18.7, pp. 1025–1033.

Sylwestrak, E. L. and A. Ghosh (2012). "Elfn1 regulates target-specific release probability at CA1-interneuron synapses". In: *Science* 338.6106, pp. 536–540.

Szabadics, J., C. Varga, G. Molnár, S. Oláh, P. Barzó, and G. Tamás (2006). "Excitatory effect of GABAergic axo-axonic cells in cortical microcircuits". In: *Science* 311.5758, pp. 233–235.

Tanay, A. and A. Sebé-Pedrós (2021). "Evolutionary cell type mapping with single-cell genomics". In: *Trends in genetics* 37.10, pp. 919–932.

Tang, F., C. Barbacioru, Y. Wang, E. Nordman, C. Lee, N. Xu, X. Wang, J. Bodeau, B. B. Tuch, A. Siddiqui, et al. (2009). "mRNA-Seq whole-transcriptome analysis of a single cell". In: *Nature methods* 6.5, pp. 377–382.

Tasic, B. (2018). "Single cell transcriptomics in neuroscience: cell classification and beyond". In: *Current opinion in neurobiology* 50, pp. 242–249.

Tasic, B., V. Menon, T. N. Nguyen, T. K. Kim, T. Jarsky, Z. Yao, B. Levi, L. T. Gray, S. A. Sorensen, T. Dolbeare, et al. (2016). "Adult mouse cortical cell taxonomy revealed by single cell transcriptomics". In: *Nature neuroscience* 19.2, pp. 335–346.

Tasic, B., Z. Yao, L. T. Graybuck, K. A. Smith, T. N. Nguyen, D. Bertagnolli, J. Goldy, E. Garren, M. N. Economo, S. Viswanathan, et al. (2018). "Shared and distinct transcriptomic cell types across neocortical areas". In: *Nature* 563.7729, pp. 72–78.

Team, R. C. (2021). *R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2012*.

team, T. pandas development (Feb. 2020). *pandas-dev/pandas: Pandas*. Version latest.

Thomson, J. A. (1925). *Concerning evolution*. Yale University Press.

Tomioka, N. H., H. Yasuda, H. Miyamoto, M. Hatayama, N. Morimura, Y. Matsumoto, T. Suzuki, M. Odagawa, Y. S. Odaka, Y. Iwayama, et al. (2014). "Elfn1 recruits presynaptic mGluR7 in trans and its loss results in seizures". In: *Nature communications* 5.1, pp. 1–16.

Tosches, M. A. (2021a). "Different origins for similar brain circuits". In: *Science* 371.6530, pp. 676–677.

Tosches, M. A. (2021b). "From cell types to an integrated understanding of brain evolution: the case of the cerebral cortex". In: *Annual Review of Cell and Developmental Biology* 37, pp. 495–517.

Tosches, M. A. and G. Laurent (2019). "Evolution of neuronal identity in the cerebral cortex". In: *Current opinion in neurobiology* 56, pp. 199–208.

Tosches, M. A., T. M. Yamawaki, R. K. Naumann, A. A. Jacobi, G. Tushev, and G. Laurent (2018). "Evolution of pallium, hippocampus, and cortical cell types revealed by single-cell transcriptomics in reptiles". In: *Science* 360.6391, pp. 881–888.

Traag, V. A., L. Waltman, and N. J. Van Eck (2019). "From Louvain to Leiden: guaranteeing well-connected communities". In: *Scientific reports* 9.1, p. 5233.

Tremblay, R., S. Lee, and B. Rudy (2016). "GABAergic interneurons in the neocortex: from cellular properties to circuits". In: *Neuron* 91.2, pp. 260–292.

Tripathy, S. J., L. Toker, B. Li, C.-L. Crichlow, D. Tebaykin, B. O. Mancarci, and P. Pavlidis (2017). "Transcriptomic correlates of neuron electrophysiological diversity". In: *PLoS computational biology* 13.10, e1005814.

Tsodyks, M., K. Pawelzik, and H. Markram (1998). "Neural networks with dynamic synapses". In: *Neural computation* 10.4, pp. 821–835.

Tsodyks, M. V., W. E. Skaggs, T. J. Sejnowski, and B. L. McNaughton (1997). "Paradoxical effects of external modulation of inhibitory interneurons". In: *Journal of neuroscience* 17.11, pp. 4382–4388.

Udakis, M., V. Pedrosa, S. E. Chamberlain, C. Clopath, and J. R. Mellor (2020). "Interneuron-specific plasticity at parvalbumin and somatostatin inhibitory synapses onto CA1 pyramidal neurons shapes hippocampal output". In: *Nature communications* 11.1, pp. 1–17.

Ulinski, P. (2007). "Visual cortex of turtles". In: *Evolution of Nervous Systems* 2, pp. 195–203.

Uller, T., A. P. Moczek, R. A. Watson, P. M. Brakefield, and K. N. Laland (2018). "Developmental bias and evolution: A regulatory network perspective". In: *Genetics* 209.4, pp. 949–966.

Urban-Ciecko, J. and A. L. Barth (2016). "Somatostatin-expressing neurons in cortical networks". In: *Nature Reviews Neuroscience* 17.7, pp. 401–409.

Vallejos, C. A., D. Risso, A. Scialdone, S. Dudoit, and J. C. Marioni (2017). "Normalizing single-cell RNA sequencing data: challenges and opportunities". In: *Nature methods* 14.6, pp. 565–571.

Van der Maaten, L. and G. Hinton (2008). "Visualizing data using t-SNE." In: *Journal of machine learning research* 9.11.

Van Rossum, G. and F. L. Drake (2009). *Python 3 Reference Manual*. Scotts Valley, CA: CreateSpace. ISBN: 1441412697.

Van Vreeswijk, C. and H. Sompolinsky (1996). "Chaos in neuronal networks with balanced excitatory and inhibitory activity". In: *Science* 274.5293, pp. 1724–1726.

vanRossum, G. (1995). "Python reference manual". In: *Department of Computer Science [CS]* R 9525.

Virshup, I., S. Rybakov, F. J. Theis, P. Angerer, and F. A. Wolf (2021). "anndata: Annotated data". In: *bioRxiv*.

Virtanen, P. et al. (2020). “SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python”. In: *Nature Methods* 17, pp. 261–272.

Vogels, T. P. and L. F. Abbott (2005). “Signal propagation and logic gating in networks of integrate-and-fire neurons”. In: *Journal of neuroscience* 25.46, pp. 10786–10795.

Vogels, T. P., H. Sprekeler, F. Zenke, C. Clopath, and W. Gerstner (2011). “Inhibitory plasticity balances excitation and inhibition in sensory pathways and memory networks”. In: *Science* 334.6062, pp. 1569–1573.

Vyas, S., M. D. Golub, D. Sussillo, and K. V. Shenoy (2020). “Computation through neural population dynamics”. In: *Annual review of neuroscience* 43, pp. 249–275.

Wamsley, B. and G. Fishell (2017). “Genetic and activity-dependent mechanisms underlying interneuron diversity”. In: *Nature Reviews Neuroscience* 18.5, pp. 299–309.

Wang, J., D. Narain, E. A. Hosseini, and M. Jazayeri (2018). “Flexible timing by temporal scaling of cortical responses”. In: *Nature neuroscience* 21.1, pp. 102–110.

Wang, X., Y. He, Q. Zhang, X. Ren, and Z. Zhang (2021). “Direct comparative analyses of 10X genomics chromium and smart-seq2”. In: *Genomics, proteomics & bioinformatics* 19.2, pp. 253–266.

Wang, Y., M. Toledo-Rodriguez, A. Gupta, C. Wu, G. Silberberg, J. Luo, and H. Markram (2004). “Anatomical, physiological and molecular properties of Martinotti cells in the somatosensory cortex of the juvenile rat”. In: *The Journal of physiology* 561.1, pp. 65–90.

Waskom, M. L. (2021). “seaborn: statistical data visualization”. In: *Journal of Open Source Software* 6.60, p. 3021.

Wehr, M. and A. M. Zador (2003). “Balanced inhibition underlies tuning and sharpens spike timing in auditory cortex”. In: *Nature* 426.6965, pp. 442–446.

Wei, J.-R., Z.-Z. Hao, C. Xu, M. Huang, L. Tang, N. Xu, R. Liu, Y. Shen, S. A. Teichmann, Z. Miao, et al. (2022). “Identification of visual cortex cell types and species differences using single-cell RNA sequencing”. In: *Nature Communications* 13.1, p. 6902.

Werbos, P. J. (1990). “Backpropagation through time: what it does and how to do it”. In: *Proceedings of the IEEE* 78.10, pp. 1550–1560.

Wester, J. C., V. Mahadevan, C. T. Rhodes, D. Calvigioni, S. Venkatesh, D. Maric, S. Hunt, X. Yuan, Y. Zhang, T. J. Petros, et al. (2019). “Neocortical projection neurons instruct inhibitory interneuron circuit development in a lineage-dependent manner”. In: *Neuron* 102.5, pp. 960–975.

Wichterle, H., D. H. Turnbull, S. Nery, G. Fishell, and A. Alvarez-Buylla (2001). “In utero fate mapping reveals distinct migratory pathways and fates of neurons born in the mammalian basal forebrain”. In: *Development*.

Williams, L. E. and A. Holtmaat (2019). “Higher-order thalamocortical inputs gate synaptic long-term potentiation via disinhibition”. In: *Neuron* 101.1, pp. 91–102.

Williams, S. R. and G. J. Stuart (2002). “Dependence of EPSP efficacy on synapse location in neocortical pyramidal neurons”. In: *Science* 295.5561, pp. 1907–1910.

Williams, S. R. and G. J. Stuart (1999). “Mechanisms and consequences of action potential burst firing in rat neocortical pyramidal neurons”. In: *The Journal of physiology* 521.Pt 2, p. 467.

Wilmes, K. A. and C. Clopath (2019). "Inhibitory microcircuits for top-down plasticity of sensory representations". In: *Nature communications* 10.1, p. 5055.

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

Wonders, C. P. and S. A. Anderson (2006). "The origin and specification of cortical interneurons". In: *Nature Reviews Neuroscience* 7.9, pp. 687–696.

Wonders, C. P., L. Taylor, J. Welagen, I. C. Mbata, J. Z. Xiang, and S. A. Anderson (2008). "A spatial bias for the origins of interneuron subgroups within the medial ganglionic eminence". In: *Developmental biology* 314.1, pp. 127–136.

Wong, K.-F. and X.-J. Wang (2006). "A recurrent network mechanism of time integration in perceptual decisions". In: *Journal of Neuroscience* 26.4, pp. 1314–1328.

Wu, S. J., E. Sevier, D. Dwivedi, G.-A. Saldi, A. Hairston, S. Yu, L. Abbott, D. H. Choi, M. Sherer, Y. Qiu, et al. (2023). "Cortical somatostatin interneuron subtypes form cell-type-specific circuits". In: *Neuron* 111.17, pp. 2675–2692.

Wu, Y. K., C. Miehl, and J. Gjorgjieva (2022). "Regulation of circuit organization and function through inhibitory synaptic plasticity". In: *Trends in Neurosciences*.

Wunderlich, T. C. and C. Pehle (2021). "Event-based backpropagation can compute exact gradients for spiking neural networks". In: *Scientific Reports* 11.1, p. 12829.

Xiang, Z., J. R. Huguenard, and D. A. Prince (1998). "Cholinergic switching within neocortical inhibitory networks". In: *Science* 281.5379, pp. 985–988.

Xu, N.-l., M. T. Harnett, S. R. Williams, D. Huber, D. H. O'Connor, K. Svoboda, and J. C. Magee (2012). "Nonlinear dendritic integration of sensory and motor input during an active sensing task". In: *Nature* 492.7428, pp. 247–251.

Xu, Q., I. Cobos, E. De La Cruz, J. L. Rubenstein, and S. A. Anderson (2004). "Origins of cortical interneuron subtypes". In: *Journal of Neuroscience* 24.11, pp. 2612–2622.

Xue, M., B. V. Atallah, and M. Scanziani (2014). "Equalizing excitation–inhibition ratios across visual cortical neurons". In: *Nature* 511.7511, pp. 596–600.

Yamins, D. L., H. Hong, C. F. Cadieu, E. A. Solomon, D. Seibert, and J. J. DiCarlo (2014). "Performance-optimized hierarchical models predict neural responses in higher visual cortex". In: *Proceedings of the national academy of sciences* 111.23, pp. 8619–8624.

Yang, G. R., J. D. Murray, and X.-J. Wang (2016). "A dendritic disinhibitory circuit mechanism for pathway-specific gating". In: *Nature communications* 7.1, pp. 1–14.

Yang, G. R. and X.-J. Wang (2020). "Artificial neural networks for neuroscientists: a primer". In: *Neuron* 107.6, pp. 1048–1070.

Yao, Z., C. T. van Velthoven, T. N. Nguyen, J. Goldy, A. E. Sedeno-Cortes, F. Baftizadeh, D. Bertagnolli, T. Casper, M. Chiang, K. Crichton, et al. (2021). "A taxonomy of transcriptomic cell types across the isocortex and hippocampal formation". In: *Cell* 184.12, pp. 3222–3241.

Yartsev, M. M. (2017). "The emperor's new wardrobe: rebalancing diversity of animal models in neuroscience research". In: *Science* 358.6362, pp. 466–469.

Yavorska, I. and M. Wehr (2016). "Somatostatin-expressing inhibitory interneurons in cortical circuits". In: *Frontiers in neural circuits* 10, p. 76.

Yizhar, O., L. E. Fenno, M. Prigge, F. Schneider, T. J. Davidson, D. J. O'shea, V. S. Sohal, I. Goshen, J. Finkelstein, J. T. Paz, et al. (2011). "Neocortical excitation/inhibition balance in information processing and social dysfunction". In: *Nature* 477.7363, pp. 171–178.

Young, H., B. Belbut, M. Baeta, and L. Petreanu (2021). "Laminar-specific cortico-cortical loops in mouse visual cortex". In: *Elife* 10, e59551.

Yu, B., Q. Zhang, L. Lin, X. Zhou, W. Ma, S. Wen, C. Li, W. Wang, Q. Wu, X. Wang, et al. (2023). "Molecular and cellular evolution of the amygdala across species analyzed by single-nucleus transcriptome profiling". In: *Cell Discovery* 9.1, p. 19.

Yu, J., H. Hu, A. Agmon, and K. Svoboda (2019). "Recruitment of GABAergic interneurons in the barrel cortex during active tactile behavior". In: *Neuron* 104.2, pp. 412–427.

Yuste, R., M. J. Gutnick, D. Saar, K. R. Delaney, and D. W. Tank (1994). "Ca2+ accumulations in dendrites of neocortical pyramidal neurons: an apical band and evidence for two functional compartments". In: *Neuron* 13.1, pp. 23–43.

Zeisel, A., A. B. Muñoz-Manchado, S. Codeluppi, P. Lönnerberg, G. La Manno, A. Juréus, S. Marques, H. Munguba, L. He, C. Betsholtz, et al. (2015). "Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq". In: *Science* 347.6226, pp. 1138–1142.

Zeng, H. (2022). "What is a cell type and how to define it?" In: *Cell* 185.15, pp. 2739–2755.

Zenke, F. and S. Ganguli (2018). "Superspike: Supervised learning in multilayer spiking neural networks". In: *Neural computation* 30.6, pp. 1514–1541.

Zenke, F. and T. P. Vogels (2021). "The remarkable robustness of surrogate gradient learning for instilling complex function in spiking neural networks". In: *Neural computation* 33.4, pp. 899–925.

Zheng, G. X., J. M. Terry, P. Belgrader, P. Ryvkin, Z. W. Bent, R. Wilson, S. B. Ziraldo, T. D. Wheeler, G. P. McDermott, J. Zhu, et al. (2017). "Massively parallel digital transcriptional profiling of single cells". In: *Nature communications* 8.1, p. 14049.

Zhou, L., T. Feng, S. Xu, F. Gao, T. T. Lam, Q. Wang, T. Wu, H. Huang, L. Zhan, L. Li, et al. (2022). "ggmsa: a visual exploration tool for multiple sequence alignment and associated data". In: *Briefings in Bioinformatics* 23.4, bbac222.

Zhu, F., S. Elnozahy, J. Lawlor, and K. V. Kuchibhotla (2023). "The cholinergic basal forebrain provides a parallel channel for state-dependent sensory signaling to auditory cortex". In: *Nature neuroscience*, pp. 1–10.

Zhu, J. J. (2000). "Maturation of layer 5 neocortical pyramidal neurons: amplifying salient layer 1 and layer 4 inputs by Ca2+ action potentials in adult rat tuft dendrites". In: *The Journal of physiology* 526.3, pp. 571–587.

Zipser, D. and R. A. Andersen (1988). "A back-propagation programmed network that simulates response properties of a subset of posterior parietal neurons". In: *Nature* 331.6158, pp. 679–684.

Znamenskiy, P., M.-H. Kim, D. R. Muir, M. F. Iacaruso, S. B. Hofer, and T. D. Mrsic-Flogel (2018). "Functional selectivity and specific connectivity of inhibitory neurons in primary visual cortex". In: *bioRxiv*, p. 294835.

A

Supporting information to Chapter 4

A.1 Network & optimization parameters

A.2 Mathematical analysis of a simplified model

We performed a mathematical analysis of a simplified network to better understand the following results of our spiking network simulations:

1. A compartment-specific balance requires PV $\rightarrow$ SST inhibition, but no other IN $\rightarrow$ IN connectivity (Fig. 4.5).

2. Higher interneuron rates require less IN specialization, i.e., individual interneurons often inhibit both PC compartments (Fig A.2).

The simplified model consists of a population of principal cells (PC) and two populations of interneurons that we will refer to as parvalbumin (PV)-positive and as somatostatin (SST)-positive cells. The population activity of the PCs is represented by somatic activity e and dendritic activity b. The interneuron activities are represented by firing rates p and s. The four activity variables e, b, p, s are best thought of as deviations of the respective activity from baseline. The activity variables can hence be both positive and negative (ignoring saturation effects that arise when the baseline is very low, see below).

For our analysis, we make the following assumptions: (1) somatic input linearly increases somatic activity e, (2) dendritic input linearly increases dendritic activity b, which is in turn assumed to be independent of somatic input/activity (note that the latter assumption deviates from a BAC-firing mechanism Larkum et al., 1999, but is necessary to obtain a linear model), (3) the activities p, s of the interneuron populations increase linearly with their input, and (4) short-term plasticity is characterized by a single, static parameter (see below). Because we are interested only in qualitative statements, the analysis is done in terms of unitless variables.

Table A.1: Parameter values related to network simulation

Symbol	Value	Unit	Description
N_E	400	-	Number of exc. neurons
N_I	100	-	Number of inh. neurons
E_L	-70	mV	reversal and reset potential
ϑ	-50	mV	spiking threshold
$\tau_{s/d/i}$	16 / 7 / 10	ms	time const. soma/ dend./inh. membrane
τ_r	3	ms	refractory time soma and inh.
$g_{s/d}$	1300 / 1200	pA	Coupling from dend to soma
$C_{s/d/i}$	370 /170/100	pF	Conductance of soma/dend./inh.
$\tau_{s/d,w}$	100 / 30	ms	Time const. adaptation soma/dend.
b_s	-200	pA	Spike-triggered adaptation (soma)
a_d	-13	nS	Voltage-driven adaptation (dend)
c_d	2600	pA	Coupling soma to dend.
E_d	-38	mV	position dend. nonlinearity
D_d	6	mV	steepness of dend. nonlinearity
$\mu_{s/d/i}$	400 / -300 /-100	pA	mean background input soma/dend./inh.
$\sigma_{s/d/i}$	450 / 450 / 400	pA	sd background input
τ_{bg}	2	ms	time const. background input
τ_{syn}	5	ms	time const. synapses
τ_u	100	ms	time const. facilitation
τ_R	100	ms	time const. depression
F	0.1	-	facilitation jump

The model describes the dynamics of the four activity variables e, b, p, s:

$$\dot{e} = -e - w^{ep}p + E^e(t), \tag{A.1}$$

$$\dot{b} = -b - w^{bs}s + E^b(t), \tag{A.2}$$

$$\dot{p} = -p + \alpha w^{pe}e + (1-\alpha)w^{pe}b - w^{ps}s, \tag{A.3}$$

$$\dot{s} = -s + \beta w^{se}e + (1-\beta)w^{se}b - w^{sp}p. \tag{A.4}$$

Here, the synaptic weight from population y to x is modeled with a non-negative weight w^{xy} ($x, y \in \{e, p, s\}$; e: PCs, p: PV INs, s: SST INs). The central tenet of this simplified model is that somatic and dendritic activity both generate characteristic spike patterns in PCs—such as events and bursts—which are selectively transmitted by synapses because of short-term plasticity. The parameters $\alpha, \beta \in [0, 1]$ describe the short-term plasticity of the PC→PV and PC→SST synapses, respectively. $\alpha, \beta = 1$ corresponds to synapses that only transmit somatic activity. If somatic activity generates events and dendritic activity generates bursts, this would require "perfectly depressing" synapses, i.e., synapses that transmit only the first spike of a burst. $\alpha, \beta = 0$ corresponds to synapses that only transmit dendritic activity. For the case where dendritic activity generates bursts, this requires "perfectly facilitating" synapses that ignore individual spikes and transmit only bursts. We assumed that the projections of the interneurons are specialized, i.e., that PV interneurons inhibit the soma and SST interneurons inhibit the dendrite. We will abandon this assumption in Section A.2. We also excluded inhibitory recurrence within the two populations (PV → PV, SST → SST), because these connections would only change the effective time constant of the respective activation variable.

Table A.2: Parameter values related to optimization

Symbol	Value / Init. Distribution	Dimensions	Description
U	$\mathcal{U}(0.1, .25)$	$N_E \times N_I$	Initial release prob.
$W^{E\to I}$	$\mathcal{N}(0, 1/N_E)$	$N_E \times N_I$	Exc. to Inh. weight
$W^{I\to I}$	$\mathcal{N}(0, 1/N_I)$	$N_I \times N_I$	Inh. to Inh. weight
$W^{I\to D}$	$\mathcal{N}(0, 0.2/N_I)$	$N_I \times 1$	Inh. to Exc. Dend. weight
$W^{I\to S}$	$\mathcal{N}(0, 0.2/N_I)$	$N_I \times 1$	Inh. to Exc. Soma weight
-	1e-3	-	learning rate for weights
-	4e-3	-	learning rate for U
β	10	-	Slope spiking derivative
-	1.0	-	Gradient (absolute value) clipping

The somata and dendrites of the PCs receive time-varying external inputs $E^e(t)$ and $E^b(t)$, respectively. All activity variables follow leaky dynamics.

The dynamical system can be written as $\dot{r} = Wr + I$, where the vector r contains the activation variables $r = (e, b, p, s)^T$, I contains the external inputs $I = (E^e, E^b, 0, 0)^T$, and W is the matrix of effective connectivity strengths

$$W = \begin{pmatrix} -1 & 0 & -w^{ep} & 0 \\ 0 & -1 & 0 & -w^{bs} \\ \alpha w^{pe} & (1-\alpha)w^{pe} & -1 & -w^{ps} \\ \beta w^{se} & (1-\beta)w^{se} & -w^{sp} & -1 \end{pmatrix}. \tag{A.5}$$

Assuming that the time constant of the network is sufficiently short to adiabatically follow the input currents, we can consider the steady state by setting $\dot{r} = 0$ and solving for r:

$$Wr + I = 0 \implies r = -W^{-1}I. \tag{A.6}$$

Influence of IN→IN connections on compartment-specific E/I balance

In the steady state Eq. (A.6), the IN rates are equal to

$$p = -[W^{-1}]_{31}E^e - [W^{-1}]_{32}E^b, \tag{A.7}$$

$$s = -[W^{-1}]_{41}E^e - [W^{-1}]_{42}E^b. \tag{A.8}$$

Here, $[W^{-1}]_{ij}$ refers to the element in row i and column j of the matrix W^{-1}. Assuming that the interneurons specialize by inhibiting a single compartment, a necessary (and, up to scaling, sufficient) condition for compartment-specific balance is that the PV rate p is proportional to the external input targeting the soma and independent of the input targeting the dendrite. Similarly, the SST rate should be proportional to the external input targeting the dendrite and independent of the input targeting the soma. By Eq. (A.7), a compartment-specific balance

hence requires $[W^{-1}]_{32} = 0$ and $[W^{-1}]_{41} = 0$. Computing these matrix entries yields:

$$[W^{-1}]_{32} \propto w^{pe}(1-\alpha) - w^{ps}w^{se}(1-\beta)) = 0 \tag{A.9}$$

$$[W^{-1}]_{41} \propto -w^{pe}w^{sp}\alpha + w^{se}\beta = 0\,. \tag{A.10}$$

These equations have a simple interpretation. Each of the two terms in $[W^{-1}]_{32}$ represents a pathway by which dendritic activity reaches the PV interneurons. The first term quantifies how much dendritic activity reaches PV interneurons via the direct excitatory PC→PV projection, the second represents corresponding feedforward inhibition via the PC→SST→PV pathway. If these two pathways cancel, PV activity is independent of dendritic activity. Similarly, the two pathways in $[W^{-1}]_{41}$ that transmit somatic activity to the SST need to cancel.

What is the role of short-term plasticity? For illustration, let us first consider the limiting case of "perfect" synaptic depression ($\alpha = 1$). Perfectly depressing PC → PV synapses would imply that the PV interneurons only receive somatic activity from the PCs via the direct PC→PV pathway. The condition (A.9) then reduces to

$$[W^{-1}]_{32} \propto -w^{ps}w^{se}(1-\beta) = 0\,, \tag{A.11}$$

i.e., dendritic activity should not reach PV interneurons via the indirect PC→SST→PV pathway, because this would render PV activity dependent on dendritic activity. Because dendritic activity need to be transmitted to the SST interneurons to reach an E/I balance in the dendrite, this implies that the SST→PV connection should be absent.

"Perfect" synaptic depression ($\alpha, \beta = 1$) or facilitation ($\alpha, \beta = 0$) are hard to implement, certainly by a Markram-Tsodyks model in the presence of background activity. However, the effect of imperfect depression in the PC→PV connection ($\alpha < 1$) can be compensated by feedforward inhibition along the PC→SST→PV pathway. Similarly, imperfect PC→SST facilitation picks up somatic activity, which can then be cancelled by feedforward inhibition via the PC→PV→SST pathway. The role of IN→IN synapses is therefore to complement "imperfect" short-term plasticity in decoding compartment-specific inputs.

The observation that PV→SST connections are the most important IN→IN connections in our model results from "imperfect" facilitation in the excitatory synapses onto SST interneurons. Because events occur more frequently than bursts, the excess excitation they trigger in SST interneurons needs to be actively cancelled via the PV→SST pathway. The converse SST→PV connection is less critical because bursts are comparatively rare, such that their transmission via PC→PV synapses causes only minor disturbances of the compartment-specific E/I balance.

Influence of IN baseline firing rates on interneuron specialization

The previous analysis assumed that interneurons were specialized to inhibit a single compartment. When should we expect specialization in the first place? We can investigate this

question by extending the simplified model by inhibition from all INs onto all PC compartments:

$$W = \begin{pmatrix} -1 & 0 & -w^{ep} & -w^{es} \\ 0 & -1 & -w^{bp} & -w^{bs} \\ \alpha w^{pe} & (1-\alpha)w^{pe} & -1 & -w^{ps} \\ \beta w^{se} & (1-\beta)w^{se} & -w^{sp} & -1 \end{pmatrix}. \tag{A.12}$$

A compartment-specific balance now requires external input to be cancelled by the inhibition from both interneurons:

$$E^e = w^{ep}p + w^{es}s, \tag{A.13}$$

$$E^b = w^{dp}p + w^{ds}s. \tag{A.14}$$

Without additional constraints, this system has an infinite number of solutions, i.e., weight configurations that achieve a compartment-specific balance. However, the simple constraint of low baseline firing rates of the interneurons collapses the solution space to the specialized one ($w^{es} = w^{dp} = 0$), for the following reason.

The activity variables p, s represent deviations of the interneuron firing rates from the baseline. If the baseline is sufficiently high, these deviations can be both positive and negative. In that case, inhibition from one interneuron class can be cancelled by disinhibition from the other interneuron class. PV and SST interneurons are then both free to respond to both somatic and dendritic activity as long as the weighted sums of the inhibition and disinhibition they provide to PC somata and dendrites mirror the excitatory input to those compartments. There are many ways of doing so.

For low baseline firing rates, disinhibition is no longer available because negative deviations from baseline are limited by the fact that activities cannot be negative. For illustration, let us consider the case where both PV and SST neurons have zero baseline activity. By definition, the excitatory signals $E^{e/b}$ are zero for baseline activity because they represent deviations from baseline. Therefore, we can assume that there exists a moment where the input to one PC compartment is zero, while the input to the other compartment is positive (e.g., $E^e = 0$, $E^b > 0$). By Eq. (A.13) and because all weights and rates must be positive, $w^{ep}p + w^{es}s = 0$ implies that $w^{ep} = 0$ or $p = 0$ and $w^{es} = 0$ or $s = 0$. At least one weight has to be non-zero (otherwise balancing the soma is impossible), and at least one rate has to be non-zero (otherwise balancing the dendrite is impossible). Without loss of generality, we can conclude that $w^{ep} > 0, p = 0, w^{es} = 0$, and $s > 0$: The PC soma is only inhibited by the PV neuron. Analogously, the existence of a moment when $E^b = 0$ but $E^e > 0$ implies that $w^{ds} > 0, w^{dp} = 0$, meaning that the PC dendrite is only inhibited by the SST neuron. If the baseline activity is low but not strictly zero, this saturation argument still hold if the variations in the firing rates that are required to balance the external input are larger than the baseline activity. Low baseline firing rates therefore imply interneuron specialization because they prevent inhibition and disinhibition from non-specialized neurons to cancel.

A.3 Supplementary figures

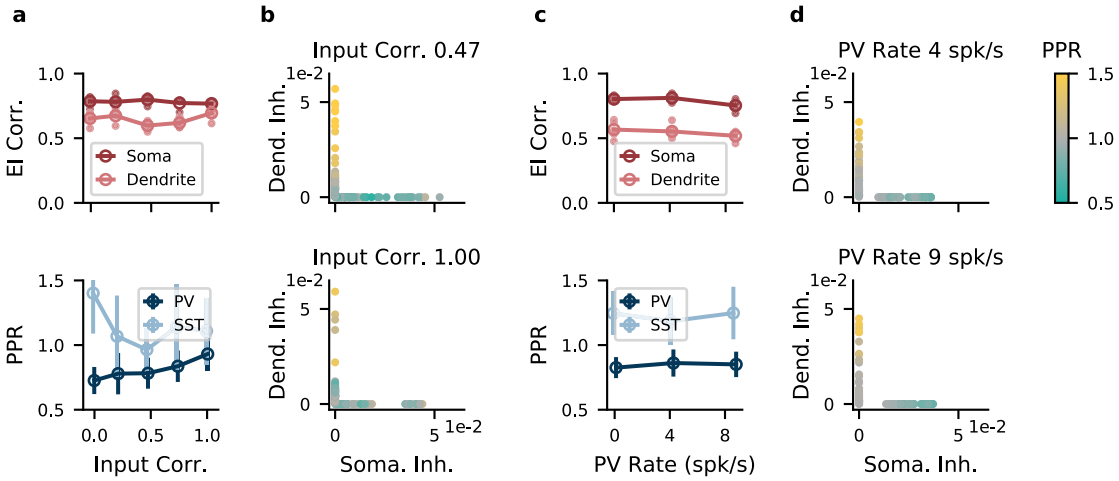

Figure A.1: Non-overlapping interneuron populations achieve compartment-specific inhibition for a range of input statistics. (a) Top, performance as measured by compartment-specific correlation between excitation and inhibition of networks trained on different correlations between compartment-specific excitatory inputs. Open circles, mean over 5 batches of 8 stimuli with random amplitudes (see Methods). Small, filled circles, individual batches. Here and in the other panels, the interneurons were assigned to inhibit only the soma or only the dendrites. Bottom, interneuron specialization, as measured by Paired Pulse Ratio (PPR), decreases with input correlations. Error bars denote sd over IN populations. **(b)** Strength of somatic and dendritic inhibition from individual INs. Top, medium input correlation (0.47); bottom, high input correlation (1.00). Color indicates PPR. **c)** Top, as **a** but as function of minimum PV rate. Bottom, interneuron specialization, as measured by Paired Pulse Ratio (PPR), is not influenced by the minimum PV rate. **(d)** Strength of somatic and dendritic inhibition from individual INs. Top, medium PV rate (4 spk/s); bottom, high PV rate (9 spk/s).

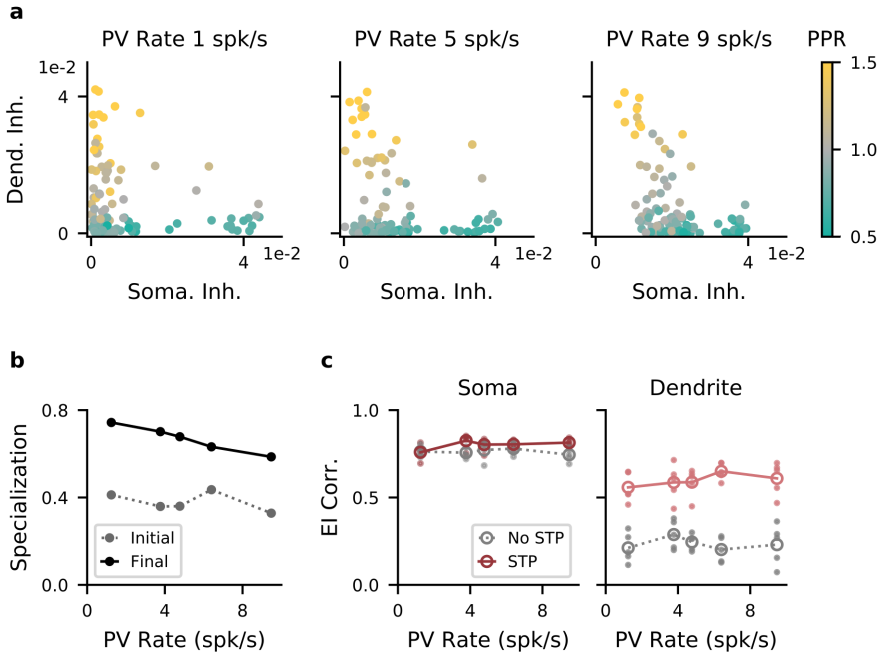

Figure A.2: Higher baseline PV rates decrease the need for interneuron specialization. **(a)** Strength of somatic and dendritic inhibition from individual INs. Left, middle, right: network optimized with a baseline PV rate of 1 (low), 5 (medium), and 9 spk/s (high), respectively. **(b)** Specialization of IN→E weights. If each IN targets either soma or dendrites, the specialization is 1 (see Methods). Grey: specialization of the initial, random network; black: specialization after optimization. **(c)** Left, the correlation between excitation and inhibition as a function of minimum PV rate. Red: networks with optimized short-term plasticity. Gray: Networks without short-term plasticity. Open circles, mean over 5 batches of 8 stimuli with random amplitudes. Small-filled circles, individual batches.

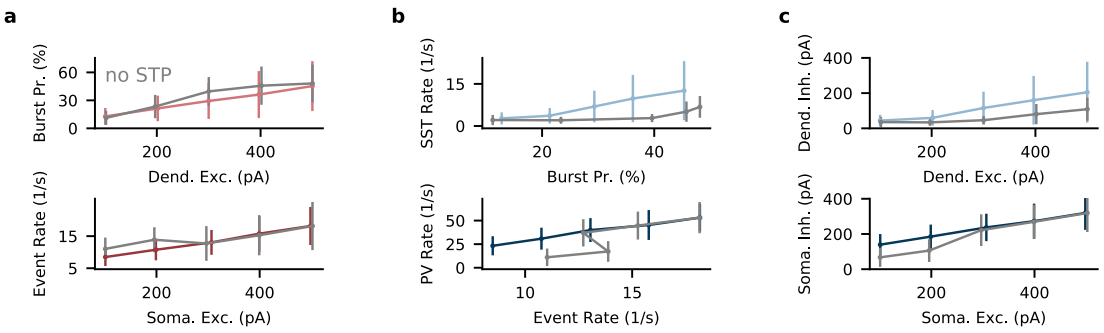

Figure A.3: Networks without short-term plasticity fail to achieve a dendrite-specific E/I balance. **(a)** PCs use a multiplexed neural code both in the presence (colours) and absence (grey) of short-term plasticity in their efferents. Top: Excitatory input to PC dendrites increases burst probability. Bottom: Excitatory input to PC somata increases event rate. **(b)** Top: SST rate increases with burst probability only when SSTs receive short-term plastic input. Bottom: PV rate increases with PC events but absent short-term plasticity only for intermediate and high event rates. **(c)** Top: dendrite-specific inhibition requires short-term plasticity. Bottom: soma-specific inhibition requires short-term plasticity only for small somatic input.

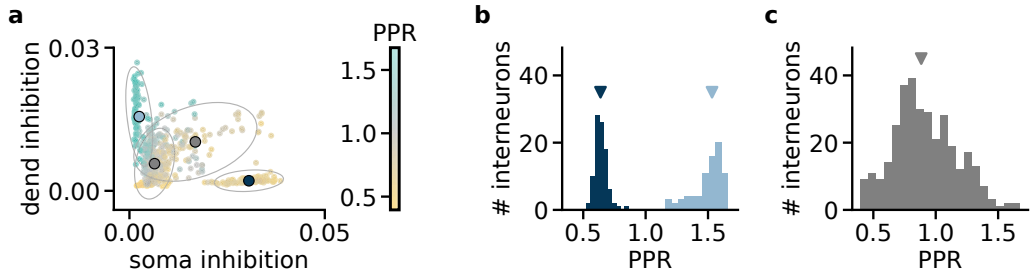

Figure A.4: Networks with heterogeneous IN→PC connections contain PV and SST classes, but also unspecific interneurons. (a) IN→PC weights after optimization in networks where INs can connect to each PC soma (and dendrite) with a unique strength. Shown are the per-IN weights averaged over all PCs. A Gaussian mixture model identified 4 clusters: a PV and a SST cluster and 2 unspecific clusters. Dots show means, ellipses show 95% density. The PV and SST clusters contain 19% and 15% of the interneurons, respectively. The unspecific clusters with small and large weights contain 52% and 14% of interneurons, respectively. **(b)** PV and SST interneurons receive depressing and facilitating inputs, respectively, as measured by the average paired-pulse ratio (PPR) computed over all presynaptic PCs. Arrows indicate means. **(c)** As (b), but for interneurons of the two unspecific clusters. Unspecific interneurons do not receive a particular type of short-term plastic input.

B

Supporting information to Chapter 6

B.1 Replication of Bugeon et al.

We validated our analysis pipeline by replicating the relevant results from Bugeon et al. Bugeon et al., 2022 on their data. These data consist of in vivo neural activity and in situ gene expression of neurons from layers 1-3 of mouse primary visual cortex (*Mus musculus* VISp). Expression data was limited to a panel of 72 genes previously selected to identify interneuron subtypes Qian et al., 2020. The data also contain behavioural variables (e.g., running speed) that assign each time point to a "behavioural state". Bugeon et al. distinguished three possible states: running (distinguished by a positive running speed), stationary desynchronized (zero running speed and little neural oscillations), and stationary synchronised (zero running speed and prominent neural oscillations). A neuron's state modulation was defined as the normalised difference between its average activity during the most and least active state:

$$\frac{\text{running activity} - \text{synchronized activity}}{\text{running activity} + \text{synchronized activity}} \tag{B.1}$$

Neurons that are more active during running compared to baseline will therefore have a positive state modulation. In contrast, neurons that are less active during running will have a negative state modulation.

We selected high-quality cells following the criteria from Bugeon et al. (Bugeon et al., 2022, see Methods), resulting in the same number of 1,065 inhibitory interneurons reported in their work. These interneurons are hierarchically distributed across 5 subclasses and 35 subtypes. State modulation could be computed for 872 neurons recorded during both running and synchronized states. As previously reported, visualising the neural activity during these states suggested differential state modulation between both cell classes and subtypes (Fig. B.1a). We then computed each neuron's state modulation based on its time-averaged activity (Equation (B.1)). Consistent with ref. Bugeon et al., 2022, state modulation was negative for Pvalb (Pearson correlation -0.13), small and positive for Sst and Lamp5 (both 0.02), and strongly positive for Vip and Sncg interneurons (0.13 and 0.21, respectively) (Fig. B.1b).

Next, we aimed to replicate the correlation between the first transcriptomic principal component (tPC1) and the state modulation. To compute tPC1, we first normalised and log-transformed the cell-by-gene expression matrix to correct for differences in sequencing depth and to stabilise the gene-count variances. We then applied principal component analysis to the transformed matrix. State modulation could indeed be predicted from tPC1 projections, both for subtypes (Fig. B.1c, leave-one-out $R^2 = 0.19$) and for individual neurons ($R^2 = 0.17$). The predictive power of tPC1 is to some extent driven by between-subclass differences in gene expression Bomkamp et al., 2019 since it is relatively modest within individual subclasses (Fig. B.2). Additional tPCs contained less information regarding state modulation: the second-best tPC (tPC29) achieved an R^2 of 0.10 and explained only 0.9% of the variance, and tPC2 did not predict state modulation at all ($R^2 = -0.01$, Fig. B.3). Together, the first 30 tPCs improved upon tPC1 ($R^2 = 0.60$, 76.2% of total variance).

Finally, we verified the correlation between state modulation and cholinergic receptor expression that might reflect a mechanistic link between state modulation and transcriptome Bugeon et al., 2022. Since cholinergic receptor expression was not measured for the in vivo recorded neurons (the 72 gene panel did not include these receptors), its relationship with state modulation can only be tested using external expression values. Following ref. Bugeon et al., 2022, we obtained these values from the publicly available data of Tasic et al. Tasic et al., 2018. We preprocessed the raw count data like the Bugeon et al. expression matrix and selected the 35 upper-layer subtypes present in the in vivo data. We then computed the average receptor expression of each subtype and compared this with its average state modulation. Linear regression showed that the expression of 5 out of 15 cholinergic receptors (or receptor subunits) could predict state modulation (Figs. B.1d, B.11). These consist of the 4 receptors shown by Bugeon et al. (Chrm3,4 and Chrna4,5, their Fig. 6b) and an additional nicotinic receptor (Chrna3).

We found one qualitative difference with previous results (Fig. 5c in ref. Bugeon et al., 2022), namely a clustering of tPC1 scores into two groups corresponding to developmental origin Lim et al., 2018a. This was caused by the log transformation used here but not in the original analyses (Fig. B.3). The log transformation is a widely used preprocessing step in the analysis of count data Luecken and Theis, 2019; Ahlmann-Eltze and Huber, 2023. However, other transformations are also possible (see, e.g., Hafemeister and Satija, 2019; Lause et al., 2021). Here, it had only a minor effect on the quantitative relationship between tPC1 and state modulation (Fig. B.3).

B.2 Network analysis

The cholinergic receptor densities in our simulations were chosen consistently with the transcriptomic and activity data, but other choices are also possible, of course. We therefore investigated the effect of varying receptor densities using mathematical analyses. In particular, we asked for which receptor densities the cholinergic effect might be different from that in the mouse. For example: does the lack of inhibitory receptors in human Pvalb cells imply that these cells are not inhibited during cholinergic modulation? And does the expression of inhibitory receptors by turtle Sst cells imply that these cells are actually inhibited?

In our simulations, all neurons receive net-positive inputs. Under these conditions, the network model contains only one nonlinearity: the rectification of dendritic activity that reaches the soma. The rectification is piecewise linear: if the dendrites are excited, the dendrites influence the soma ($w_{ed} = 1$); if the dendrites are inhibited, the dendrite remains inactive and decouple from the soma ($w_{ed} = 0$). The network dynamics are, therefore, governed by one of two connectivity matrices that only differ in the entry w_{ed}. Otherwise, the dynamics are linear:

$$\dot{\mathbf{r}} = -\mathbf{r} + W(w_{ed})\mathbf{r} + \mathbf{I}_0 + \mathbf{m}. \tag{B.2}$$

Here, $\mathbf{m}$ is the vector modelling cholinergic modulation, and $\mathbf{I}_0$ is the external input. For a given somato-dendritic coupling w_{ed} and cholinergic modulation $\mathbf{m}$, the steady state rates are found by solving $\dot{\mathbf{r}} = 0$:

$$\mathbf{r}(w_{ed}, \mathbf{m}) = [I - W(w_{ed})]^{-1}[\mathbf{I}_0 + \mathbf{m}] = A(w_{ed})[\mathbf{I}_0 + \mathbf{m}], \tag{B.3}$$

where we defined $A(w_{ed}) = [I - W(w_{ed})]^{-1}$ as the matrix that maps inputs to steady-state rates:

$$A(w_{ed}) = \begin{pmatrix} 1 - w_{ee} & w_{ed} & -w_{ep} & 0 & 0 \\ w_{de} & 1 & 0 & -w_{ds} & 0 \\ w_{pe} & 0 & 1 + w_{pp} & -w_{ps} & 0 \\ w_{se} & 0 & 0 & 1 & -w_{sv} \\ w_{ve} & 0 & 0 & -w_{vs} & 1 \end{pmatrix}^{-1} \qquad w_{ab} \geq 0 \;\; \forall a, b. \tag{B.4}$$

Below, we will compute the relevant entries of $A(w_{ed})$ up to its determinant, which is positive and therefore does not affect the entries' signs.

We use equation (B.3) to compute the modulatory effect on the network activity as the difference between the rates with and without modulation. We will consider the cases in which cholinergic modulation activates the dendrite that was silent without modulation (off→on) or inactivates the dendrite that was activated without ACh (on→off). The other two cases (on→on, off→on) can be derived analogously.

First, consider the case that modulation switches the dendrites off, as for the mouse and human circuits. The resulting change in network activity equals:

$$\begin{aligned} \Delta\mathbf{r} &= \mathbf{r}(0, \mathbf{m}) - \mathbf{r}(1, \mathbf{0}) && \text{(B.5)} \\ &= A(0)[\mathbf{I}_0 + \mathbf{m}] - A(1)[\mathbf{I}_0 + \mathbf{0}] && \text{(B.6)} \\ &= A(0)\mathbf{m} + [A(0) - A(1)]\mathbf{I}_0. && \text{(B.7)} \end{aligned}$$

In case that modulation switches the dendrites on, as for the turtle circuit, the resulting change in network activity equals:

$$\Delta\mathbf{r} = A(1)\mathbf{m} + [A(1) - A(0)]\mathbf{I}_0. \tag{B.8}$$

The first term in these equations is a linear combination of the receptor densities $\mathbf{m} = (m_e, m_d, m_p, m_s, m_v)$, describing how the cholinergic modulation of individual populations propagates through the network. In our model, pyramidal cells do not express cholinergic receptors ($m_e = m_d = 0$), such that the cholinergic effect is a linear combination of only the interneuron receptor densities. The second term in Eq. (B.8) is independent of the precise modulation and describes how the background input $\mathbf{I}_0$ propagates through the network with and without activated dendrites. Since this term is small, we ignore it in the following derivations, but it is shown in Fig. 6.5.

First, consider the cholinergic effect on Pvalb cells, which equals:

$$\Delta r_p = A_{pp} m_p + A_{ps} m_s + A_{pv} m_v. \tag{B.9}$$

Substituting the entries of $A(0)$ gives:

$$\begin{aligned} \Delta r_p &= A_{pp} m_p - A_{ps}(w_{sv} \cdot m_v - m_s) && \text{(B.10)} \\ &= (1 - w_{ee}) \cdot (1 - w_{sv} \cdot w_{vs}) \cdot m_p + && \text{(B.11)} \\ &\quad w_{sv} \cdot (1 - w_{ee})(w_{ps} \cdot m_v - m_s). && \text{(B.12)} \end{aligned}$$

An analogous equation describes Δr_s (see below). The lines in Figure 6.5 show $\Delta r_p = 0$ and $\Delta r_s = 0$ as a function of the Pvalb and Sst receptor densities m_p and m_s, for a fixed Vip density m_v. These boundaries delineate domains of positive and negative modulation of Pvalb and Sst interneurons. Gamma-aminobutyric acid

So does the absence of inhibitory ACh receptors in human Pvalb cells (Fig. 6.4; $m_p = 0$) imply that these cells will not be inhibited? Equation Eq. (B.10) shows that these cells will still be inhibited indirectly under the condition that:

$$\Delta r_p = w_{sv} \cdot (1 - w_{ee})(w_{ps} \cdot m_v - m_s) < 0. \tag{B.13}$$

The coefficient $w_{sv} \cdot (1 - w_{ee})$ is positive, assuming the recurrent connectivity is not very strong ($w_{ee} \leq 1$). Inhibition of Pvalb cells is then equivalent to

$$\Delta r_p < 0 \iff w_{ps} \cdot m_v - m_s < 0. \tag{B.14}$$

According to the expression data, Sst and Vip cells are positively modulated ($m_s, m_v > 0$). Therefore, Pvalb cells will be inhibited by ACh as long as the inhibition from Ssts is stronger than the disinhibition from Vips. Under these conditions, the limited Chrm4 expression by human Pvalb cells is compatible with their inhibition.

Let us next consider the differential expression of inhibitory ACh receptors in Sst interneurons in turtles versus mammals. Intuitively, this is expected to cause a cholinergic suppression of Sst cells in the turtle, in contrast to the mouse. In the model, the cholinergic effect on Sst cells equals:

$$\Delta r_s = A_{sp} m_p + A_{ss} m_s + A_{sv} m_v. \tag{B.15}$$

Gamma-aminobutyric acid Subsituting the entries from $A(1)$ gives:

$$\Delta r_s = A_{sp} m_p + A_{ss}(m_s - w_{sv} \cdot m_v) \tag{B.16}$$

$$= w_{ep} \cdot (w_{sv} \cdot w_{ve} - w_{se}) \cdot m_p \; + \tag{B.17}$$

$$[(1 - w_{ee} - w_{de})(1 + w_{pp}) + w_{ep} \cdot w_{pe}]\,(m_s - w_{sv} m_s). \tag{B.18}$$

For the turtle circuit, $m_p \leq 0$, since Pvalb cells (weakly) express inhibitory ACh receptors. The first term will therefore be negative if $w_{sv} \cdot w_{ve} - w_{se} > 0$. Further, $m_s < 0$ and $m_v > 0$, such that $m_s - w_{sv} m_v < 0$. The contribution of the second term will therefore be negative if:

$$(1 - w_{ee} - w_{de})(1 + w_{pp}) + w_{ep} \cdot w_{pe} > 0.$$

This will be the case unless recurrent excitation is very strong or the feedback loop between PCs and Pvalb cells is very weak. In summary, the expression of inhibitory ACh receptors by turtle Sst cells will indeed lead to their cholinergic inhibition, provided that the excitation onto Sst cells and the recurrent excitation are not too strong.

Software

An Anaconda *Anaconda Software Distribution* 2020 environment with the appropriate software will be provided along with the code (Table B.1).

Software	version
AnnData Virshup et al., 2021	0.8.0
Matplotlib Hunter, 2007	3.6.2
Numpy Harris et al., 2020	1.23.5
Pandas team, 2020	1.5.2
Python vanRossum, 1995	3.10.10
R Team, 2021	4.3.0
Scanpy Wolf et al., 2018	1.9.1
Scikit-learn Pedregosa et al., 2011	1.2.1
Scipy Virtanen et al., 2020	1.9.3
Seaborn Waskom, 2021	0.12.2
Seurat Hao et al., 2021	4.0
Snakemake Köster and Rahmann, 2012	7.8.2
Statsmodels Seabold and Perktold, 2010	0.13.5

Table B.1: Software versions.

B.3 Supplementary figures

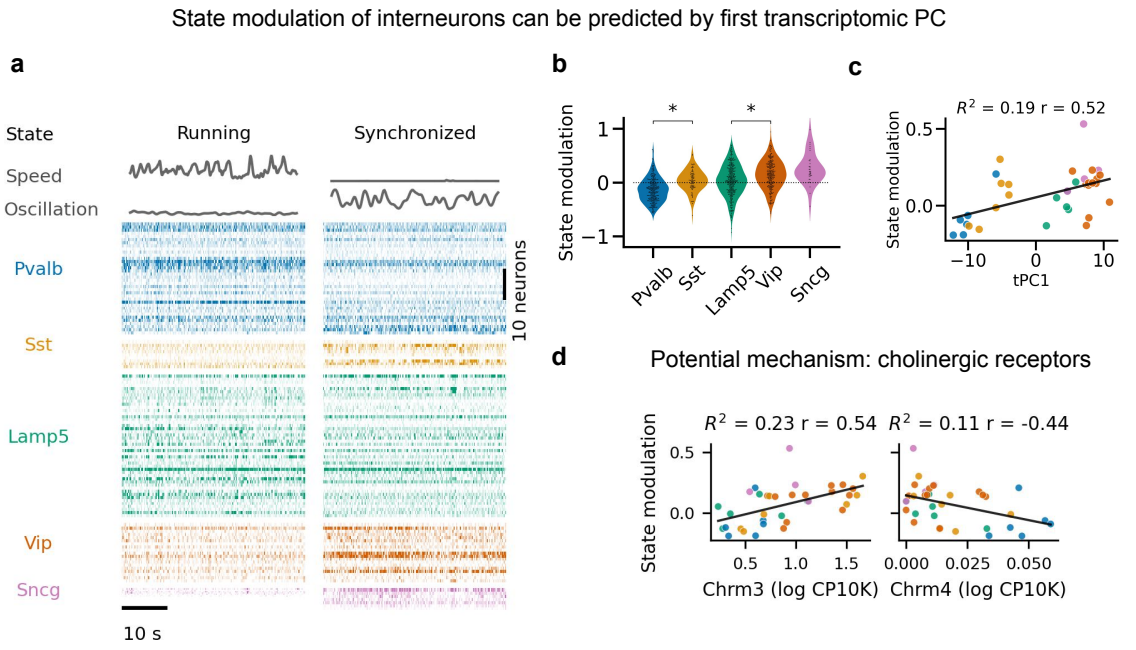

Figure B.1: Replication of previous findings from Bugeon et al. Bugeon et al., 2022 **(a)** Neural activity systematically varies with behavioural state (measured by running speed and neural oscillations, see Methods) both between and within interneuron classes of mouse primary visual cortex (VISp) L1-3. **(b)** State modulation across all sessions for $n = 872$ interneurons. Stars indicate statistically significant differences between subclasses ($p < 0.05$, Mann-Whitney U test). **(c)** The first transcriptomic principal component (tPC1) of the cell-by-gene matrix predicts state modulation of subtypes ($n = 35$); Fig. B.2a shows the relationship for individual cells. R^2, leave-one-out fraction of variance explained; r, Pearson correlation. Note the two clusters along tPC1, consisting of MGE-derived (Pvalb & Sst) and CGE-derived (Lamp5, Vip, Sncg) interneurons. **(d)** Cholinergic receptors potentially link a neuron's transcriptome and state modulation. For example, interneurons that overexpress the excitatory receptor Chrm3 are positively state-modulated ($r = 0.54$; $p = 0.0008$), those that overexpress the inhibitory cholinergic receptor Chrm4 are negatively state-modulated ($r = -0.44$, $p = 0.0075$). CP10K, counts per 10 thousand. Data and findings from Bugeon et al. Bugeon et al., 2022. Cholinergic receptor expression in (d) from Tasic et al. Tasic et al., 2018.

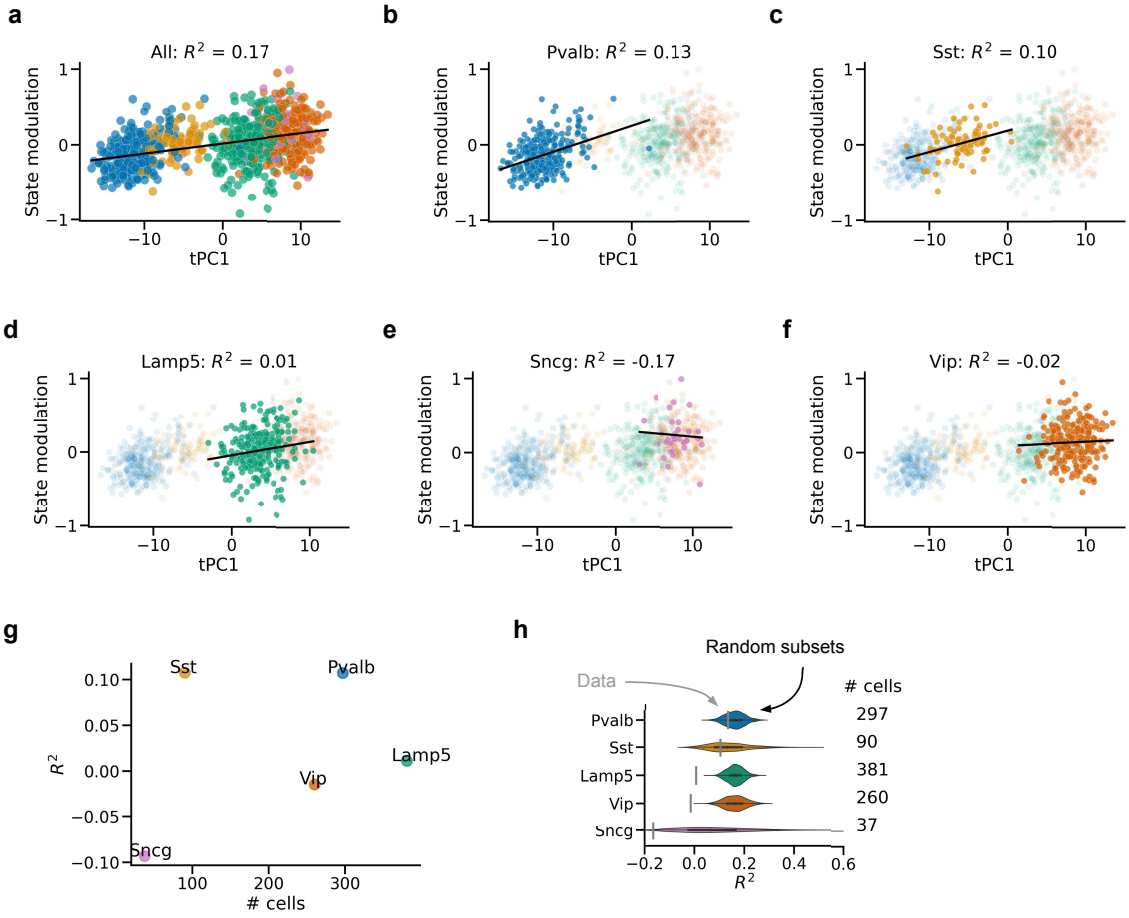

Figure B.2: Predicting state modulation from tPC1. Regression based on all cells (**a**) or cells from a single subclass (**b-f**). Predictive performance is worse for individual classes and only better than chance for Pvalb and Sst cells. The correlation between tPC1 and state modulation is therefore partially driven by between-subclass differences. However, tPC1 is still predictive of state modulation across all cells while controlling for subclass ($p = 0.003$, linear mixed model with subclass as random effect). R^2: leave-one-cell-out fraction of variance explained; $R^2 < 0$ indicates a worse fit compared to predicting the same state modulation for each cell independent of tPC1 score. **(g,h)** Poor performance for certain subclasses is not due to a smaller sample size. **(g)** Sample size is not correlated with worse performance. **(h)** Size-matched subsets of all cells outperform below-chance subclasses, except for Sst cells. Grey bars: R^2 values for each subclass. Violin plots: distribution of R^2 values for 1000 random subsets of all cells with sample size matched to the subclass. Data from Bugeon et al. Bugeon et al., 2022.

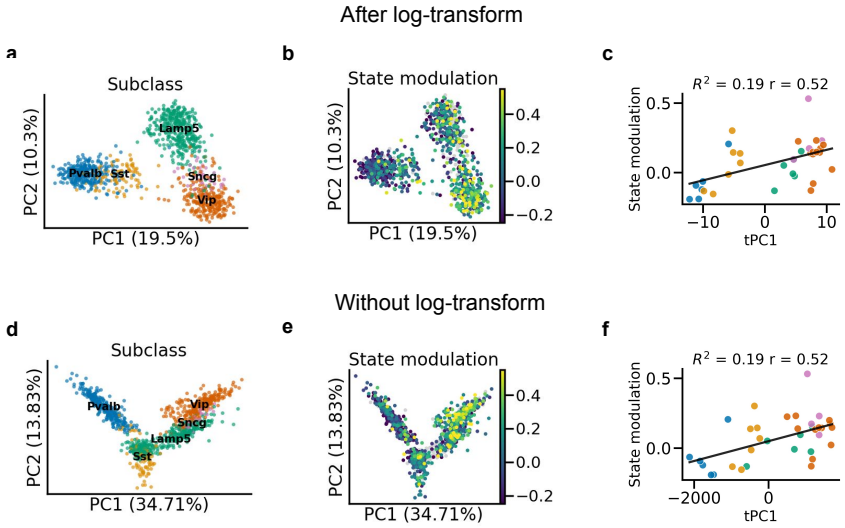

Figure B.3: Log-transformation leads to clustering by developmental origin. **(a)** First 2 transc.riptomic principal components (tPCs) of the log-transformed count RNA data. **(b)** As (a), with colour indicating state modulation. **(c)** The first transcriptomic PC (tPC1) of log-transformed data predicts state modulation, replicated from Fig. B.1c for comparison. R^2: leave-one-out fraction of variance explained, r: Pearson correlation. **(d-f)** As (a-c), but without log-transformation. Interneurons now form a continuum along tPC1, but the quantitative relationship between tPC1 and state modulation is preserved (up to 2 digits). Data from Bugeon et al. Bugeon et al., 2022.

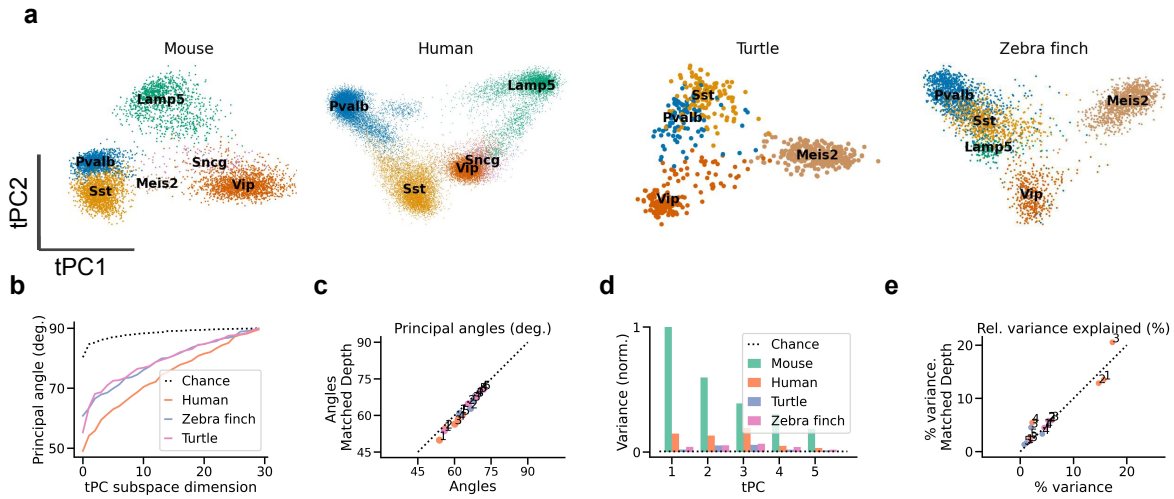

Figure B.4: Transcriptomic PCs robust to sequencing depth. **(a)** Projection of each dataset onto its first 2 transcriptomic PCs, after subsampling gene counts to the depth of the shallowest dataset (zebra finch, see Table 6.1). **(b)** Principal angles between tPC subspaces of subsampled data. **(c)** Comparison between angles of full-depth data and subsampled data. **(d,e)** As (b,c) but for variance explained. The human, turtle, and zebra finch tPC1 explain 14.0%, 1.2%, and 4.2% of the variance explained by mouse tPC1, respectively. Data from refs. Tasic et al., 2018 (mouse), Bakken et al., 2021 (human), Tosches et al., 2018 (turtle), Colquitt et al., 2021 (zebra finch).

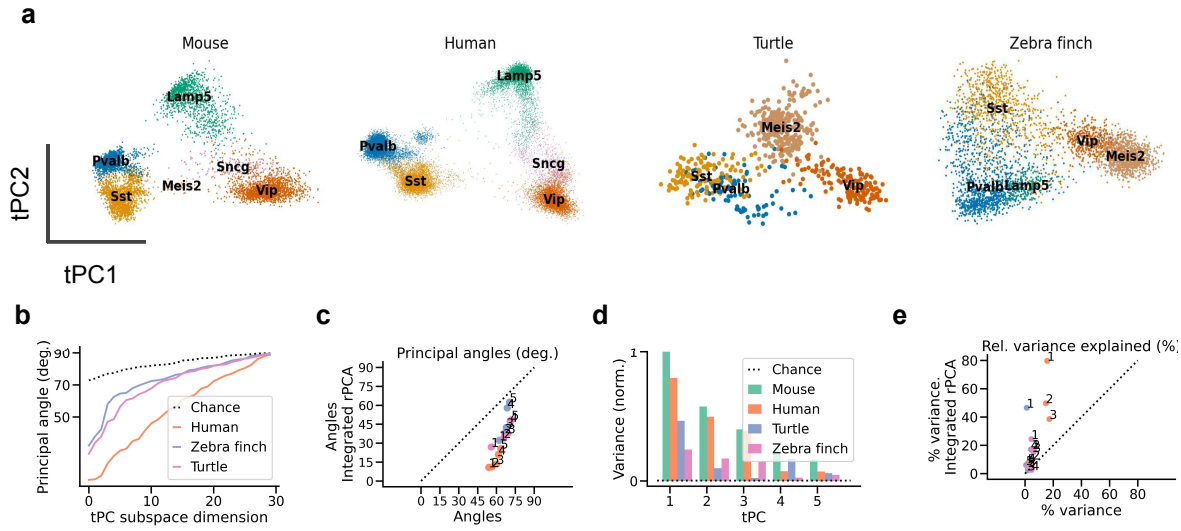

Figure B.5: Computational integration increases similarity of mouse and human data. **(a)** Projection of each dataset onto its first 2 transcriptomic PCs, after computational integration. Mouse and human datasets show increased similarity, but turtle cells no longer cluster by cell type (colour). **(b)** Principal angles between tPC subspaces computed after integration. **(c)** Comparison between angles computed without integration. Integration increased the similarity of all datasets, especially of the human data. **(d,e)** As (b,c) but for variance explained. The human, turtle, and zebra finch tPC1 explain 79.8%, 46.6%, and 24.4% of the variance explained by mouse tPC1, respectively. Data from refs. Tasic et al., 2018 (mouse), Bakken et al., 2021 (human), Tosches et al., 2018 (turtle), Colquitt et al., 2021 (zebra finch).

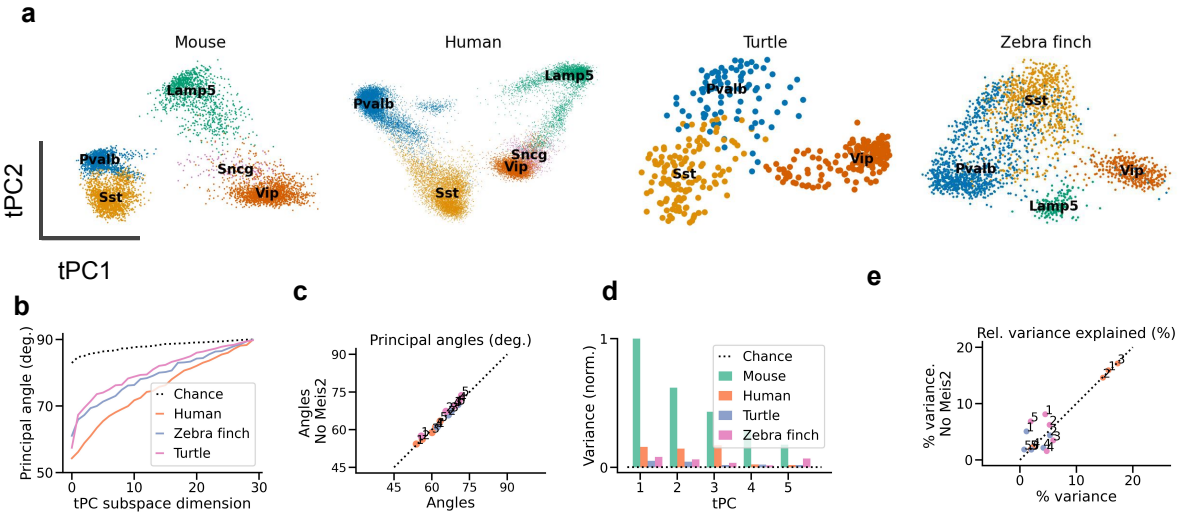

Figure B.6: Differences in tPCs not due to Meis2 cells. **(a)** Projection of each dataset onto its first 2 transcriptomic PCs, after removing Meis2 cells. **(b)** Principal angles between tPC subspaces computed without Meis2 cells. **(c)** Comparison between angles computed on all cells vs. cells without Meis2 population. **(d,e)** As (b,c) but for variance explained. The human, turtle, and zebra finch tPC1 explain 15.7%, 5.1%, and 8.1% of the variance explained by mouse tPC1, respectively. Data from refs. Tasic et al., 2018 (mouse), Bakken et al., 2021 (human), Tosches et al., 2018 (turtle), Colquitt et al., 2021 (zebra finch).

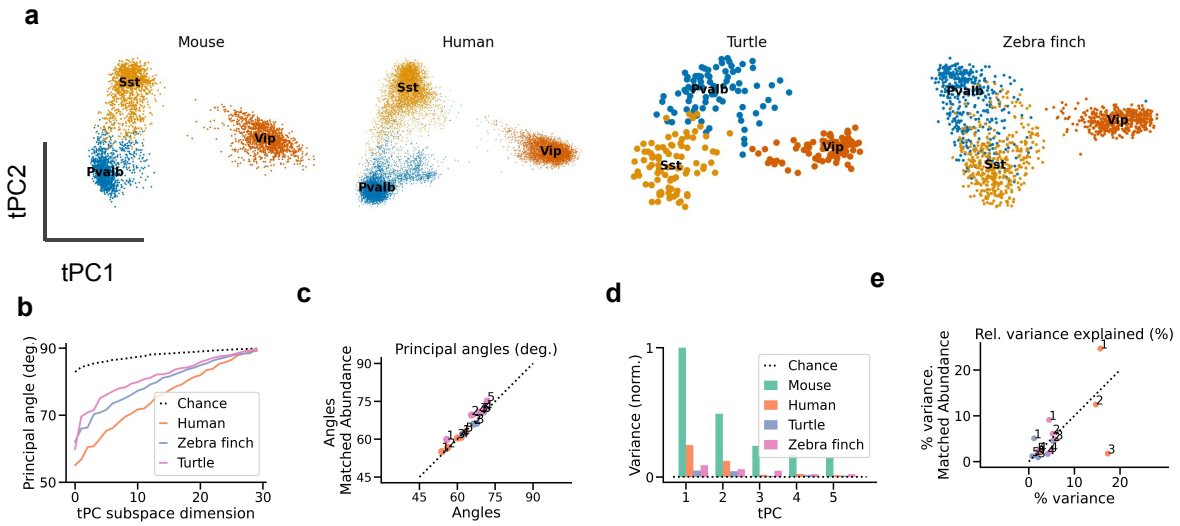

Figure B.7: Differences in tPCs not only due to cell type abundance. **(a)** Projection of each dataset onto its first 2 transcriptomic PCs, after matching cell type abundances (Fig. 6.2h). **(b)** Principal angles between tPC subspaces. **(c)** Comparison between angles computed on all cells vs. cells after matching frequencies. **(d,e)** As (b,c) but for variance explained. The human, turtle, and zebra finch tPC1 explain 24.7%, 5.1%, and 9.2% of the variance explained by mouse tPC1, respectively. Data from refs. Tasic et al., 2018 (mouse), Bakken et al., 2021 (human), Tosches et al., 2018 (turtle), Colquitt et al., 2021 (zebra finch).

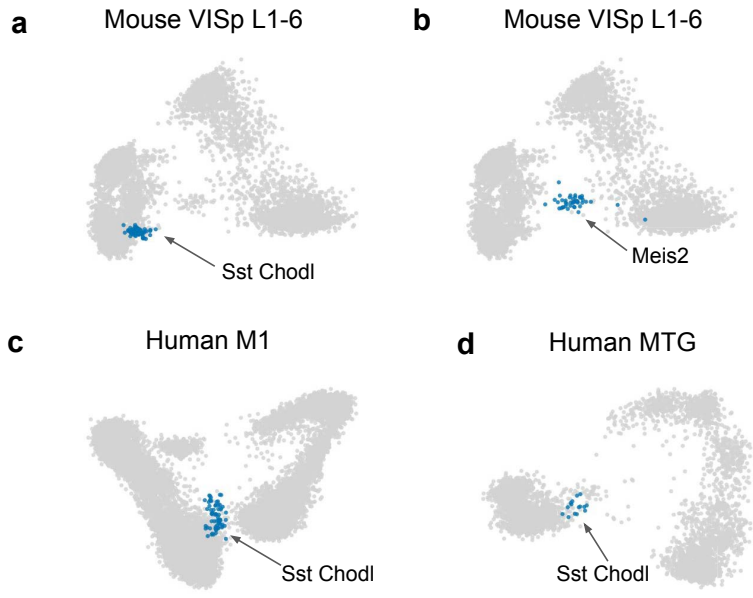

Figure B.8: Intermediate tPC1 position of Chodl and Meis2 neurons. Long range projecting Sst-Chodl **(a)** and white matter Meis2-Adamts19 cells **(b)** occupy an intermedate position along tPC1. **(c,d)** Sst Chodl neurons also have intermediate tPC1 scores in the human data. The human datasets do not contain Meis2 cells. Data from refs. Tasic et al., 2018 (a,b), Bakken et al., 2021 (c), and Hodge et al., 2019 (d).

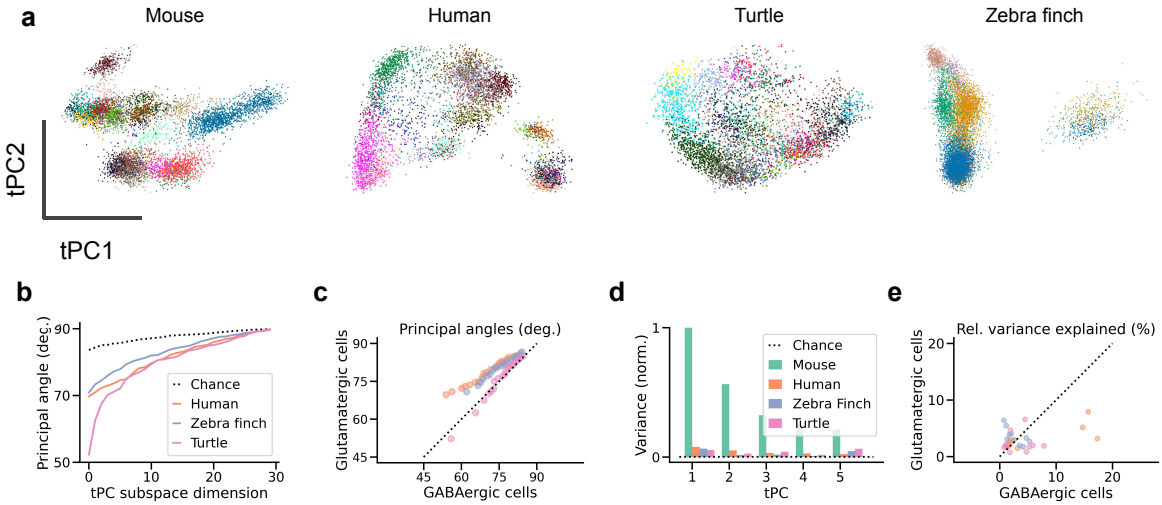

Figure B.9: Slightly larger differences between glutamatergic compared to GABAergic cells. **(a)** Projection of each dataset's glutamatergic cells onto its first 2 transcriptomic PCs. Colours indicate cluster assignment from the original publications. **(b)** Principal angles between tPC subspaces of the mouse glutamatergic cells and those of other datasets. **(c)** Comparison of angles between GABAergic subspaces and angles between glutamatergic subspaces. Only a few of the turtle angles (pink) are smaller for the glutamatergic cells. **(d,e)** As (b,c) but for variance explained. The human, turtle, and zebra finch tPC1 explain 7.9%, 5.5%, and 6.6% of the variance explained by mouse tPC1, respectively. Data from refs. Tasic et al., 2018 (mouse), Bakken et al., 2021 (human), Tosches et al., 2018 (turtle), Colquitt et al., 2021 (zebra finch).

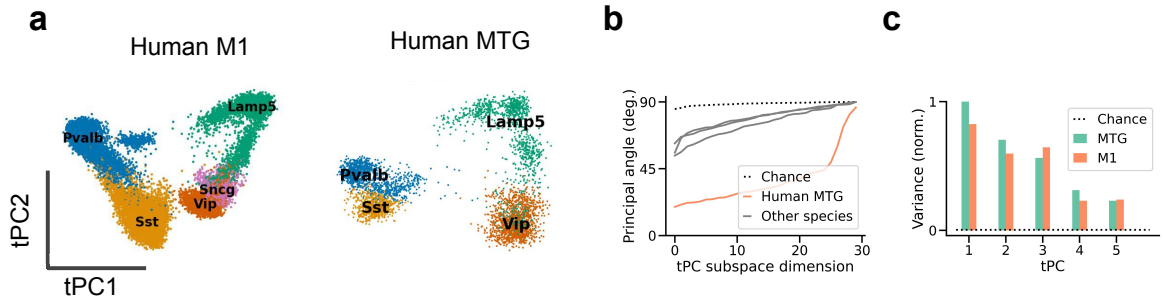

Figure B.10: Small differences in PCs of human datasets. **(a)** Projection of human datasets onto their first 2 tPCs. **(b)** Quantification of tPC similarity using principal angles between tPC subspaces of M1 and MTG data. **(c)** Quantification by variance explained in MTG data. The M1 tPC1 explains 82.6% of the MTG variance explained by the MTG tPC1. Data from refs. Bakken et al., 2021 and Hodge et al., 2019.

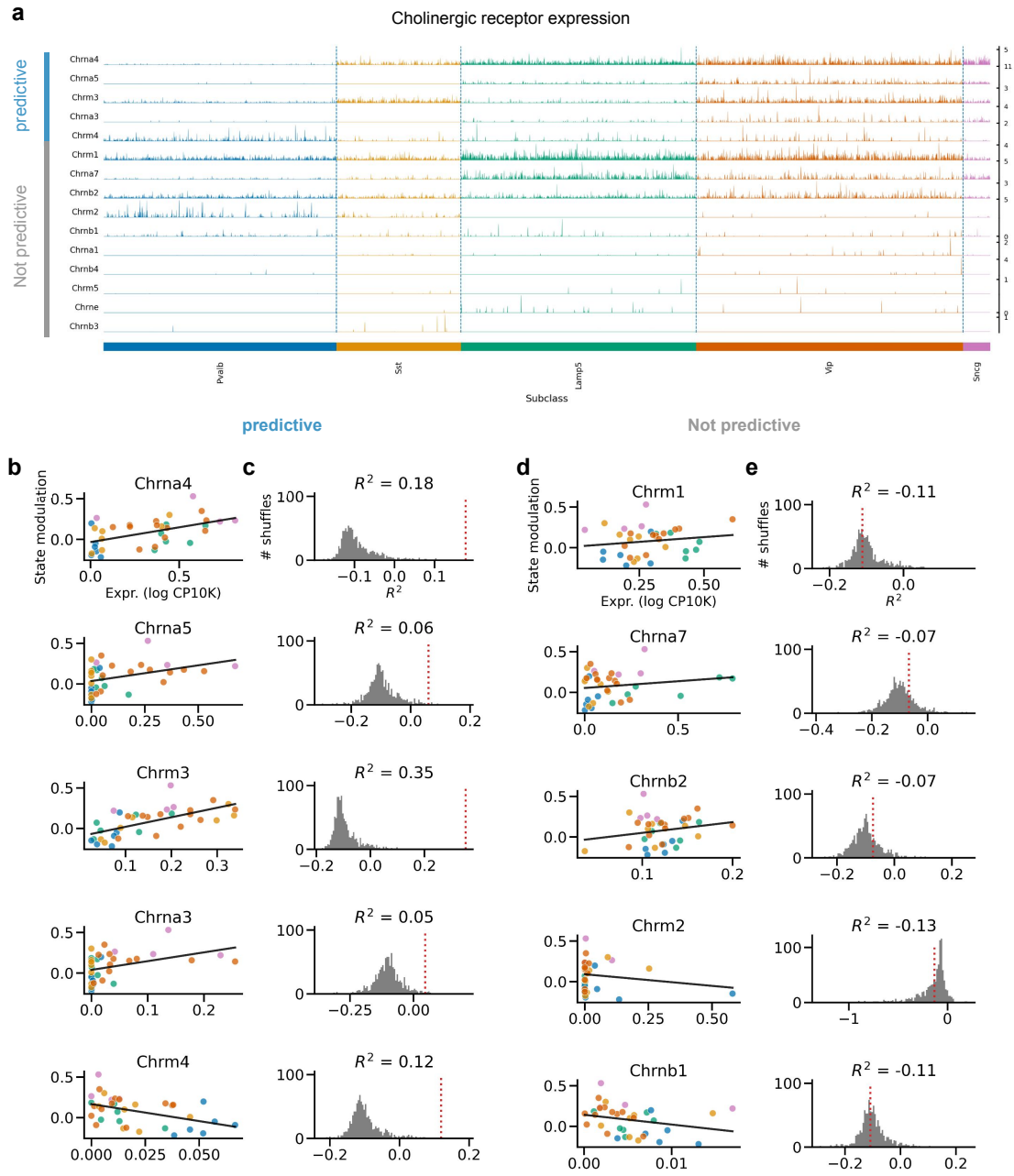

Figure B.11: Predicting state modulation from cholinergic receptor expression. (a) Tracks plot of cholinergic receptor (subunit) expression. The first 5 receptors predict state modulation; the remaining 10 do not (see b-e). Predictive and unpredictable receptors are independently sorted by expression based on expression levels. Shown are all receptors with an expression of at least 1 count per 10K. **(b)** Relationship between state modulation and log expression of receptors that are predictive of state modulation (1000 permutations, $p < 0.05$). **(c)** Grey: Null distribution of leave-one-out R^2 estimated by linear regression after permuting expression levels. Red: R^2 without permutation. **(d,e)** As (b,c) but for the 5 unpredictive receptors with the highest expression. Receptor expression from Tasic et al. Tasic et al., 2018; state modulation from Bugeon et al. Bugeon et al., 2022.

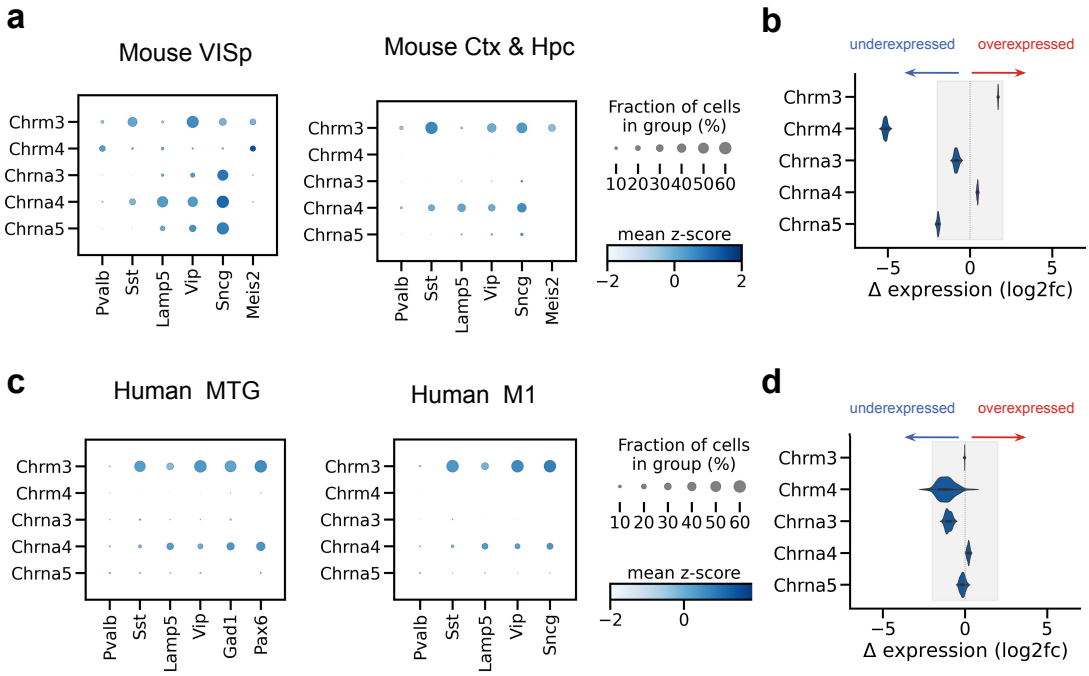

Figure B.12: Mostly small within-species differences in ACh receptor expression. (a) Dot plots showing the expression of the cholinergic receptors that predict state modulation in mouse VISp L1-3 **(b)** Log2-fold differences in expression after downsampling the VISp dataset to equal sequencing depth as the Ctx & Hpc data. Shaded area: log-fold difference of ±2, the range of most within-species differences. The exception is Chrm4, which is underexpressed in the Ctx & Hpc data compared to the VISp data. **(c,d)** As (a,b), but for human datasets. The MTG dataset was downsampled to match the M1 data. Data from refs. Tasic et al., 2018 (mouse VISp), Yao et al., 2021 (mouse Ctx & Hc), Hodge et al., 2019 (human MTG), and Bakken et al., 2021 (human M1).