Inaugural dissertation

for

obtaining the doctoral degree

of the

Combined Faculty of Mathematics, Engineering and Natural Sciences

of the

Ruprecht - Karls – University

Heidelberg

Presented by

M.Sc. Francisca Hervas Sotomayor

born in: Quito, Ecuador

Oral examination: February 2nd, 2024

Tracing the origins and evolution of the vertebrate brain: a comparative cellular study in anamniotes

Referees:

Prof. Dr. Henrik Kaessmann

Prof. Dr. Detlev Arendt

I. Summary

Vertebrates, a diverse group of organisms adapted to a range of ecological niches, rely on external cues for responding to environmental pressures. The brain plays a crucial role in processing this information and coordinating the corresponding responses. Despite this adaptation diversity, the fundamental patterning of this organ into prosencephalon, mesencephalon, and rhombencephalon appears to be conserved across all extant vertebrate species, suggesting an ancestral organization. Although scientists have studied the brain comparatively to understand its origin and evolution, most research has focused on amniotes. This narrow focus, however, has hampered a clear understanding of the origin of multiple structures that predate this lineage.

In this thesis, I address this gap by comparatively analyzing the cell composition of the brains of multiple non-amniotes. My goal was to better understand the cellular and molecular origins of this organ and its diversification through evolution. To accomplish that, the first step involved generating a comprehensive and spatially resolved transcriptomic cell type atlas for the brain of the sea lamprey, a cyclostome whose phylogenetic position allows for the inference of ancestral vertebrate traits. This endeavor was complemented by producing and integrating data for catshark, spotted gar, and lungfish; species that belong to main gnathostome lineages, enabling trait reconstruction within the vertebrate clade.

By comparing broad cell classes between these species, I discovered conserved expression profiles of transcription factors and effector genes, indicating that these classes are homologous across vertebrates. The comparative analysis between mice and lamprey atlases further discerned shared cell type families. Additionally, the identification of the main embryonic sources of telencephalic inhibitory neurons in the lamprey brain, confirmed their existence in vertebrate ancestors. The analyses also revealed key tissues and cell types that probably emerged later in evolution, after the divergence of cyclostomes and gnathostomes. For instance, the ancestral brain probably lacked cerebellar cells and oligodendrocytes (myelinating cells); the latter likely evolved in gnathostomes from an astrocyte-like cell. It seems that the ancestral glia already possessed certain elements of the molecular machinery involved in oligodendrocyte differentiation and myelin production. However, crucial genes from this machinery appeared only in gnathostomes. This indicates that the genome

duplication, which occurred in this group, played a role in the emergence of these cell types. Furthermore, it's likely that the vertebrate ancestor had a single-domain pallium homologous to the multiple-domain pallium present in tetrapods. The analyses between gnathostomes suggested that while there is a clear indication of homology among the pallial neurons at a broad scale, the existence of one-to-one homologies remains questionable.

Collectively, my research identifies the ancestral cellular configuration and molecular core of the vertebrate brain. Additionally, I provide insights into the cellular diversification that has accompanied the evolution of the clade. Notably, these findings not only address unresolved questions in comparative neuroscience but also point to new directions concerning the temporal and mechanistic dynamics behind the evolution of tissues, such as the pallium, amygdala and cerebellum. Finally, by focusing on anamniotes, these results contribute substantially to the refinement of brain evolution models in vertebrates.

II. Zusammenfassung

Vertebraten haben sich an eine Reihe ökologischer Nischen angepasst. Im Anpassungsprozess müssen sie äußere Reize verarbeiten, um auf den Umweltdruck zu reagieren. Das Gehirn spielt hier eine entscheidende Rolle bei der Verarbeitung und der Koordination von Verhaltens- und Stoffwechselreaktionen. Trotz der Vielzahl an Nischen und der draus folgenden Anpassungsvielfalt scheint die grundlegende Gliederung dieses Organs in Prosencephalon, Mesencephalon und Rhombencephalon bei allen existierenden Wirbeltierarten erhalten zu sein. Dies weist auf eine ursprüngliche Organisation hin, die vor ca. 515 – 645 Millionen Jahren entstanden ist. Um diesen Ursprung und seine Evolution zu verstehen, gab es schon eine Vielzahl vergleichender Studien. Jedoch konzentrierte sich die meiste Forschung auf Amnioten. Die Fokussierung auf die Amnioten erlaubt allerdings kein klares Verständnis des Ursprungs der Hirnstrukturen, die dieser letzten Klade vorausgingen.

In meiner Doktorarbeit gehe ich auf diese genau Wissenslücke ein, indem ich die Zellzusammensetzung der Gehirne mehrerer Nicht-Amnioten vergleichend analysiere. Mein Ziel war es, systematisch die zellulären und molekularen Ursprünge dieses Organs und seine Diversifizierung im Laufe der Evolution besser zu verstehen. Um dies zu erreichen, bestand der erste Schritt darin, einen umfassenden und räumlich aufgelösten transkriptomischen Zelltypenatlas für das Gehirn des Neunauges zu erstellen. Neunaugen sind Zyklostomen. Ihre phylogenetische Position erlaubt Rückschlüsse auf ursprüngliche Wirbeltiermerkmale. Diese Untersuchung wurde komplettiert mit Daten aus Katzenhaie, gefleckten Knochenhechte und Lungenfische; die Arten gehören zu der Hauptlinien der Gnathostomen, was eine Rekonstruktion der ursprünglichen Merkmale innerhalb der Vertebraten ermöglicht.

Durch den Vergleich von Zellklassen zwischen diesen Arten entdeckte ich konservierte Expressionsprofile von Transkriptionsfaktoren und Effektor-Genen, die auf eine Homologie eben dieser Klassen von Wirbeltieren hinweisen. Diese Analyse ergänzte ich durch den systematischen Vergleich zwischen Mäuse- und Neunaugen-Atlanten, der gemeinsame Zelltypfamilien aufgedeckt hatte. Des Weiteren identifizierte ich die primäre embryonale Region der inhibitorischen Neuronen des Telencephalons im Gehirn des Neunauges. Das macht ihre Existenz in den Wirbeltiervorfahren sehr plausibel. Umgekehrt enthüllten die dieselben Analysen auch Gewebe und Zelltypen, die wohl später in der Evolution entstanden – nach der Abspaltung von Zyklostomen und Gnathostomen. Zum Beispiel fehlten dem

ursprünglichen Gehirn wahrscheinlich Zerebellarzellen und Oligodendrozyten (myelinisierende Zellen); letztere sind vermutlich in den Gnathostomen aus einer astrozytenähnlichen Zelle entstanden. Es scheint, dass das ursprüngliche Glia bereits gewisse Teile der molekularen Maschinerie besaß, welche an der Oligodendrozytendifferenzierung und Myelinproduktion beteiligt ist. Allerdings finden sich wichtige Gene aus dieser Maschinerie nur bei den Gnathostomen. Dies deutet darauf hin, dass die Genomverdopplung, die in dieser Gruppe auftrat, eine Rolle bei der Entstehung dieser Zelltypen spielte. Darüber hinaus ist anzunehmen, dass der Wirbeltiervorfahre ein einziges Pallium besaß, das dem der Tetrapoden mit mehreren Domänen homolog ist. Die Analysen zwischen den Gnathostomen deuteten darauf hin, dass es zwar klare Indikatoren für eine allgemeine Homologie unter den pallialen Neuronen gibt, jedoch die Existenz von eins-zu-eins-Homologien fraglich bleibt.

Zusammengenommen identifizieren meine Forschungsergebnisse den ursprünglichen zellulären Kern und die molekulare Konfiguration des Wirbeltiergehirns. Überdies liefere ich Einblicke in die zelluläre Diversifizierung des Gehirns, die die Evolution der Vertebraten begleitet hat. Diese Ergebnisse bringen somit Erkenntnisse zu unbeantworteten Fragen der vergleichenden Neurowissenschaft und geben Hinweise bezüglich der zeitlichen und mechanistischen Dynamiken hinter der Evolution der Gewebe. Schließlich kann so durch die Einbeziehung von Nicht-Amnioten das Modell der Gehirnevolution bei Wirbeltieren verfeinert werden.

I dedicate this thesis to all the bright minds out there who couldn't fulfill their dream of becoming a scientist, particularly those inspiring people I have met who happened to be born in less fortunate circumstances.

IV. Acknowledgments

One of the highlights of being a scientist is the teamwork, the human component of this profession that enriches the experience. I could not have reached this point without the support of multiple people.

First, I would like to express my gratitude to my supervisor, Prof. Dr. Henrik Kaessmann, for offering me this opportunity. Over the last years, I was lucky to contribute to a groundbreaking project, and I'm thankful for the confidence he placed in me and for his support. His mentorship and stimulating discussions on science and evolution, among other topics, have been invaluable for my professional and personal development.

My appreciation extends to Dr. Mari Sepp and Dr. Francesco Lamanna for their significant involvement in my project from its beginning. Their guidance laid the foundation for my scientific development over these last years. Their continuous support and enjoyable conversations about the brain paved the way for this dissertation. Additionally, I thank Dr. A. Phillip Oel, my lab mate from the onset of my PhD journey. He has been extremely supportive, and I have enjoyed every moment I worked with him. I take so many good memories from both professional and personal experiences, and I hope our paths remain close.

During the last years of my PhD, I have had the privilege of collaborating closely with the Vertebrain group members, Dr. Kerry Lynn Gendreau and Matthias Janeschik, as well as with inspiring students and new members. They have been a formidable team and I appreciate the discussions, the validation, and the complicity. At a crucial moment in my PhD, they reminded me of my equal love for science and nature and the reason why I pursued a career in Evolutionary Biology in the first place. Additionally, I extend my thanks to them, as well as to Bastienne Zaremba and Daniel Soto Carballo, for their valuable feedback on this thesis.

I am also grateful to the European Commission for funding the initial phase of my PhD through the EvoCELL initiative. I would like to extend a special thank you to all EvoCELL members for the enriching experiences, training and discussions we had together. Particularly, I thank Dr. Francesca Stomeo, Dr. Jacob Musser, and all the remarkable fellows. This experience was an example of science at its most enjoyable.

My deepest appreciation goes to the members of my advisory committee, Prof. Dr. Detlev Arendt and Dr. Moritz Mall, for their essential guidance and support. In particular, I thank them both and their teams for providing me with lab space when most needed, as well as for the warm welcome into their groups. I also extend my thanks to Prof. Dr. Maximilian J. Telford, who, apart from providing mentorship under the EvoCELL initiative, agreed to read this dissertation and be part of my examination committee.

This project was made possible through collaborations with numerous research groups, who contributed samples and invaluable expertise. I would like to acknowledge Dr. Antón Barreiro-Iglesias and his team, who additionally hosted us during the collection of lamprey brains. Special recognition to his former student, Dr. Daniel Sobrido-Camean, for his expertise which was critical to this project. My thanks also go to Dr. Daniel M. Medeiros, Dr. David Jandzik, and the teams of Dr. Marianne E. Bronner, Dr. Sylvie Mazan, and Dr. Tae Sung Jung, whose contributions were indispensable.

Many people from the Kaessmann lab deserve recognition, yet Dr. Florent Murat deserves a special mention. His friendship and support, both professional and personal, have been invaluable during my journey. I'm more than thrilled to have the opportunity to continue my career by working with him. I'm also really thankful to all former and current members of the Kaessmann lab for their help and providing such a friendly environment to work. Special thanks go to Nils Trost, Noe Mbengue, Céline Schneider, and Julia Schmidt for their hands-on involvement and assistance. I am particularly grateful for the daily interactions with Céline, Noe, Julia, and Marta, filled with jokes, laughs, and conversations. A warm thank you to all other lab members, in particular Bassi, Leti and Margarida, for their friendship.

On the other hand, I would like to thank my Ecuadorian biologist friends, including those scattered across the globe, for believing in me over the last 17 years. Whenever I've doubted my path, they've reminded me why we keep doing this. Their passion for nature, science, and making a difference is remarkable; I can only wish to work with them again.

Additionally, I want to express my gratitude to the friends I met in Heidelberg—Gonzalo, Fernando, Alexandra, Amon, Rossen, Tico, Amandine, Akash, Cinthia, Nataly, and Fernando. Moving far from home is never easy, but sometimes life rewards you with a new family.

I'm, of course, thankful to all my family, especially my mother for encouraging me to chase my dreams while supporting me in every possible way. To my sister Daniela and cousin Ana; our frequent visits, mutual encouragement and cartoonish humor are a source of strength. In addition, I want to thank my German family, who accepted me, opened their homes, and supported me during this journey.

I'm deeply and truly thankful to my partner, Markus. His understanding, love, and complicity are impossible to explain in a few words. He has been by my side through difficulties, successes, tears, and laughter over the years. He encourages me to pursue my dreams and creates the safest space possible for me to be myself.

Thanks, gracias, yupaychani!

V. Table of Contents

VI. Table of figures

1 Introduction

Vertebrates, a diverse clade of animals that have successfully colonized various environments and ecological niches, showcase a wide range of adaptations. While it is difficult to make broad statements about selection pressures because of the wide array of niches these species inhabit, certain common patterns can still be observed. For instance, these animals face challenges associated with reproduction, predation avoidance, and foraging[1]. To effectively address these pressures, they must perceive information from their surroundings, encompassing both environmental and social cues. This information is then processed and responded to accordingly. The central nervous system, primarily the brain, plays a crucial role in acquiring, processing, and reacting to this information, ultimately shaping behavioral and metabolic responses[1,2].

Despite the diverse adaptations among vertebrates, the central nervous system has retained a general conserved patterning, indicating the action of constraints or stabilizing selection throughout the evolution of the clade. This conserved pattern is evident in the brain's fundamental division into forebrain (prosencephalon), midbrain (mesencephalon), and hindbrain (rhombencephalon)[3] (Fig. 1a). Because this "bauplan" is present in all extant vertebrates, and the genetic program governing it appears to be conserved as well, it is believed that it was already established in the last common ancestor of the entire clade[3,4] (Fig. 1b).

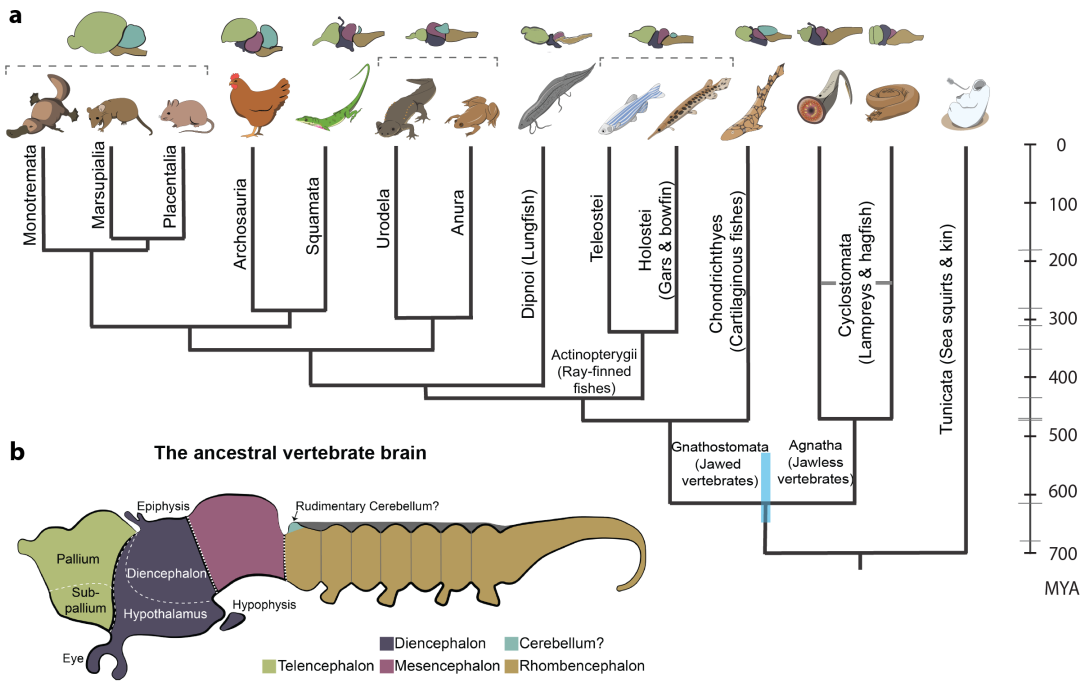


Figure 1. Vertebrate brain overview. a) Phylogenetic tree displaying major vertebrate lineages and their approximate brain anatomies; the blue bar indicates the estimated confidence interval for the divergence time of cyclostomes and gnathostomes (*adapted from Lamanna, Hervas-Sotomayor et al*[5]*, license under Creative Commons CC BY*). b) Schematic of the ancestral vertebrate brain. Note that this graphic depicts the forebrain divided in: telencephalon and diencephalon, following classical embryonic neuroanatomy (*schematic follows Sugahara et al.*[6]*).* Illustrations by Nils Trost and Marta Sánchez-Delgado.

Numerous studies have attempted to decipher further the ancestral organization of the vertebrate brain, through comprehensive neuroanatomy, connectivity, and gene expression analyses, often within a comparative framework[4,6–8]. Gene expression data, in particular, have offered convincing evidence for a vertebrate brain blueprint partially shared with other members of the chordate clade[9–12] (Fig. 2). For instance, the molecular pathways directing the brain's developmental partitioning into its four main domains seem to be preserved among vertebrates (Fig. 2b). Additionally, these data indicate the maintenance of cell proliferation and migration patterns within the vertebrate brain[1].

1.1 Developmental divisions of the vertebrate brain

During development, three primary brain divisions can be identified, each emerging as a distinctive enlargement of the neural tube (Fig. 2b). As development progresses, these divisions further subdivide and significant changes are seen, such as the hindbrain developing into distinct swellings called rhombomeres . The most rostral and dorsal rhombomere develops into the cerebellum and cerebellum-like structures in most vertebrates. Furthermore, the hindbrain also gives rise to the pons and medulla oblongata. The midbrain primarily develops into the optic tectum[1,13] (Fig. 3a). However, the most drastic changes are observed in the forebrain. Caudally, it develops into the diencephalon, which includes multiple structures such as the pretectum, thalamus, and prethalamus. Rostrally, the secondary prosencephalon emerges, encompassing the hypothalamus, retina (both sometimes identified within the diencephalon), preoptic area, and the telencephalon, which further divides into the subpallium (ventral telencephalon) and pallium (dorsal telencephalon)[14,15] (Fig. 3a-b).

1.2 The origin of vertebrates and the evolution of brain structures

The origin of vertebrates was accompanied by significant changes in the brain, as evidenced by its distinctive anatomy. Upon examining the brains of the closest relatives of this clade, such

as amphioxus (cephalochordates) and tunicates (urochordates), it is clear that they are markedly different from those of vertebrates[16]. In contrast to the domains of vertebrates, the organization of the central nervous system in cephalochordates and tunicates appears to be more uniform, making it challenging to identify homologous brain structures across chordates[1] (Fig. 2a,c).

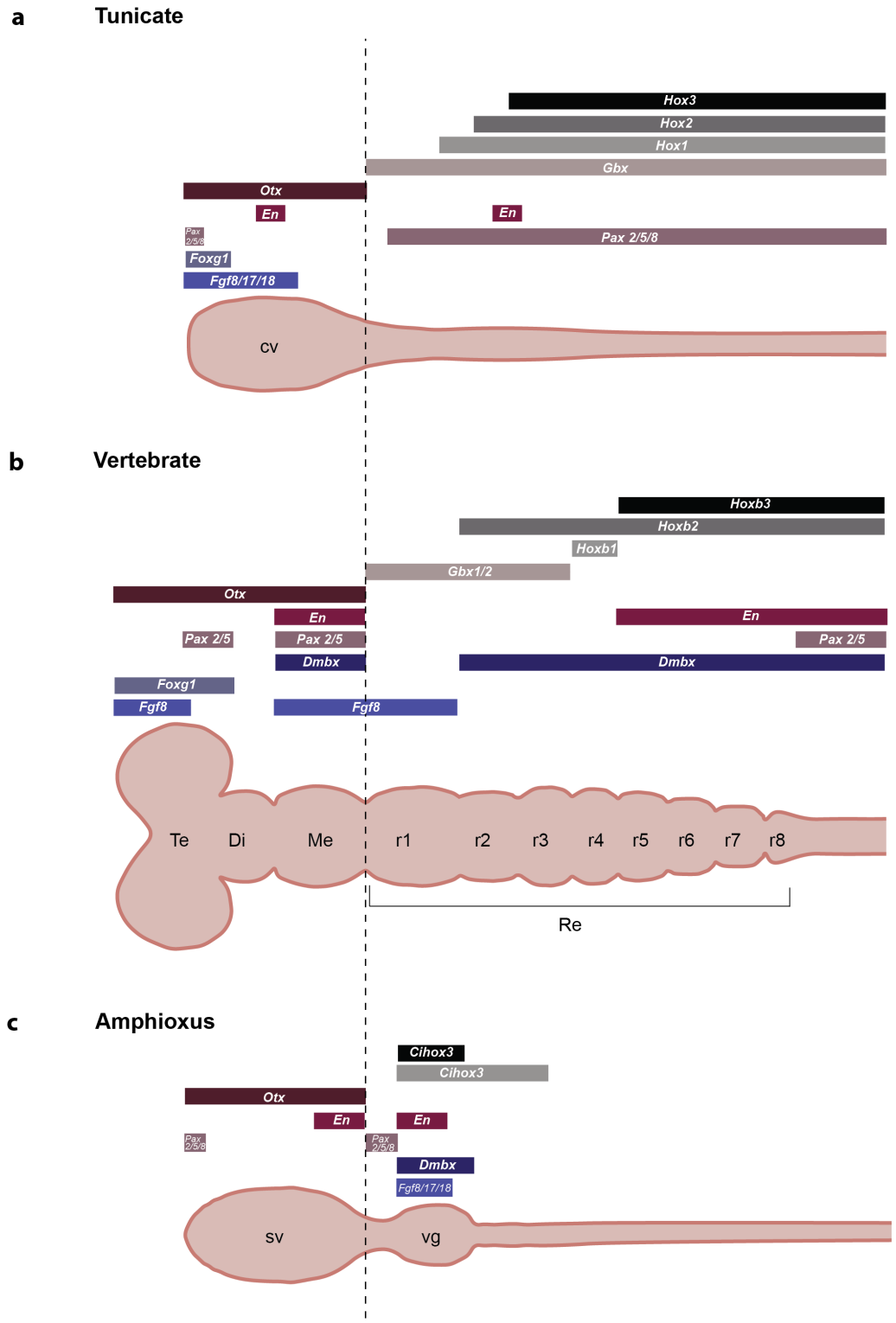


Figure 2. **Gene Expression in the central nervous system of embryonic chordates.** a) *Ciona intestinalis* CNS. b) Vertebrate brain. c) Amphioxus CNS. The dashed line indicates the caudal limit of *Otx*, which coincides with the midbrain-hindbrain boundary in the vertebrate embryonic brain. Based on these data, there is no clear evidence of telencephalon or midbrain in tunicates and amphioxus. Abbreviations: Di: diencephalon, cv: cerebral vesicle, Me: mesencephalon (midbrain), r1-8: hindbrain rhombomeres, sv: sensory vesicle, Te: telencephalon, Re: rhombencephalon (hindbrain), vg: visceral ganglion. *Redrawn from Striedter & Northcutt*[1].

Because of these marked brain differences, the extent of neural innovation following the emergence of vertebrates remains unclear. Multiple gene expression studies have illuminated this topic[17–20]. Vertebrate domains exhibit a largely conserved gene expression programming, determining their development during embryonic stages (Fig. 2b). The spatiotemporally restricted expression of pivotal transcription factors that delineate these regions is conserved throughout the clade[14,17]. However, molecular evidence coming from invertebrate chordates is not as definitive, leading to varied conclusions. Most researchers concur that all chordates possess a hindbrain (although without a cerebellum) and a forebrain predominantly composed of the retina and hypothalamus[1,20]. This suggests that the midbrain and telencephalon might be innovations of vertebrates. Still, recent data indicate that a domain positive for telencephalic markers can be found in adult amphioxus as well as in the anterior zone of the brain vesicle and placodes of tunicates[18,19].

The evolution of the midbrain and telencephalon likely correlated with the advent of pattern vision and an expanded olfactory system, as explained under the "New Head" hypothesis[21]. This hypothesis provides a comprehensive explanation for the emergence of particular embryonic tissues such as some brain domains, neural crest and placodes, believed to have played a role in the evolution of distinct structures in vertebrates that set them apart from other chordates[1,21].

1.3 Challenges in tracing the origin and evolution of the vertebrate brain

The extent to which the vertebrate brain has remained conserved is a subject of ongoing debate, primarily due to the limitations of previous comparative studies. Historically, these studies predominantly focused on whole-brain or individual structures[2], or relied heavily on a limited number of genetic markers or a narrow range of species. This issue is exacerbated by selection bias in comparative neuroscience research, which tends to favor species with conserved traits, leading to a deterministic view of evolution and strengthening anthropocentric conclusion[1,2,22].

For instance, the majority of vertebrate neuroanatomy principles have been inferred from amniotes[14,15,23], with only a few exceptions that include anamniote data[7,22,24]. These limitations make it challenging to distinguish between scenarios of convergent evolution and true conservation. The lack of appropriate phylogenetic representation, limited gene analysis, and absence of an evolutionary framework that integrates gene evolution, structural changes, and interspecific diversity further contribute to this challenge. Despite these limitations, the findings from these studies have played a crucial role in illuminating the origins of the human brain and have significantly contributed to the advancement of biomedical research[25]. Nevertheless, this approach has largely neglected to integrate an evolutionary framework, which is necessary for a comprehensive understanding of the evolution of the brain beyond a species or a clade.

Studying the deep origins of the vertebrate brain is particularly challenging, requiring identification of the degree to which the fundamental blueprint of this complex organ existed in the ancestor of all vertebrates[26]. This is of particular importance in comparative studies because identifying innovations or derived structures unique to certain lineages is impossible without first establishing homologies and inferring ancestral traits[27]. However, as mentioned previously, studying ancestral traits across all vertebrates presents multiple difficulties. An extra challenge arises from the fact that the closest chordate relatives to the vertebrate clade exhibit fundamental phenotypic differences[16] (Fig. 1a, Fig. 2). As discussed before, vertebrates possess a complex brain, many domains of which appear to be innovations[1,20]. Consequently, a suitable outgroup for reconstructing ancestral states, which arose after the split of vertebrates with other chordates, does not exist for this clade.

The first major split in the evolution of the vertebrate clade occurred around ~515–645 MYA (million years ago)[28] (Fig. 1a), resulting in two main lineages: jawless vertebrates (represented today by extant cyclostomes) and jawed vertebrates (gnathostomes). This has created a significant phylogenetic distance between cyclostomes and gnathostomes. Furthermore, within cyclostomes, there are two separate lineages (lampreys and hagfish) that not only exhibit phylogenetic divergence (median time: 455 MYA) but also have different brain morphologies[3,7]. These factors have complicated the identification of homologies both within cyclostomes and between cyclostomes and gnathostomes. Despite these limitations, the phylogenetic position of lampreys and hagfish makes studies on them crucial for understanding trait evolution[1,3,7,27,29]. However, the hagfish brain has a remarkably different morphology

compared to other vertebrates, lacking a pineal gland and exhibiting a layered dorsal telencephalon, which makes it less ideal for inferring ancestral traits for the clade[1,7]. Therefore, many fundamental principles regarding ancestral brain origins in vertebrates have been drawn from studies on lampreys.

1.4 Evolutionary insights from studies in the lamprey brain

The lamprey has been used as a subject of study in neurobiological research, with studies spanning diverse perspectives[24,30,31]. It is worth noting, however, that a significant portion, if not the majority, of the published work emphasizes specific brain tissues. When studied through an evolutionary lens, the evolution of these tissues appears to be associated with crucial moments in vertebrate history, suggesting a likely adaptive advantage. These structures include the cerebellum, telencephalon, and oligodendrocytes (myelinating glia of the central nervous system).

In general terms, the brain organization of lampreys mirrors the basic pattern seen in other vertebrates. Yet, inconsistencies emerge when exploring finer-resolution structures and cells. For instance, the presence of the cerebellum has been debated for over a century[32]. Some arguments advocate for a "rudimentary" cerebellum in the lamprey, primarily because of the presence of granule-like cells and developmental precursors[33,34]. Counterarguments highlight the lack of other defining cerebellar cell types, such as the Purkinje cells and cerebellar nuclei, and the absence of characteristic molecular patterning during development[16]. This evidence implies that a genuine cerebellum appeared later in evolution, possibly in animals with paired appendages, as indicated by fossil records[35]. Under this perspective, the granule cellular structure in the lamprey brain might correspond to a cerebellar-like circuitry in the hindbrain, dedicated to process signals from the vestibular and lateral line systems[32].

The lamprey telencephalon attracts significant attention from neuroscientists, primarily because of its significance in understanding the evolutionary origin and trajectory across vertebrate lineages. Given its critical role in cognitive and behavioral adaptations, this structure is considered a landmark of vertebrate innovation[36]. During development, the telencephalon divides into two principal domains: the dorsally specified pallium (discussed later) and the ventral sub-pallium, which further subdivides into the medial ganglionic eminence (MGE) and the caudal and lateral ganglionic eminences (CGE and LGE) (Fig. 3). In mammals, the

eminences are pivotal for neurodevelopment, particularly as primary sources of inhibitory (GABAergic) interneurons that migrate to the pallium and olfactory bulbs, respectively. This migration is critical during cortex development[1,3] (Fig. 3b). While evidence of the MGE and LGE is compelling in gnathostomes, their presumed absence in cyclostomes suggests a gnathostome lineage-specific emergence. However, the expression of MGE-associated transcription factors has been reported in cyclostome embryos[37]. Additionally, studies incorporating physiological and immunohistochemical approaches in adult lampreys indicate the existence of MGE and LGE-derived structures in the lamprey brain[38–41]. Consequently, the origin of the ganglionic eminences in vertebrates remains elusive.

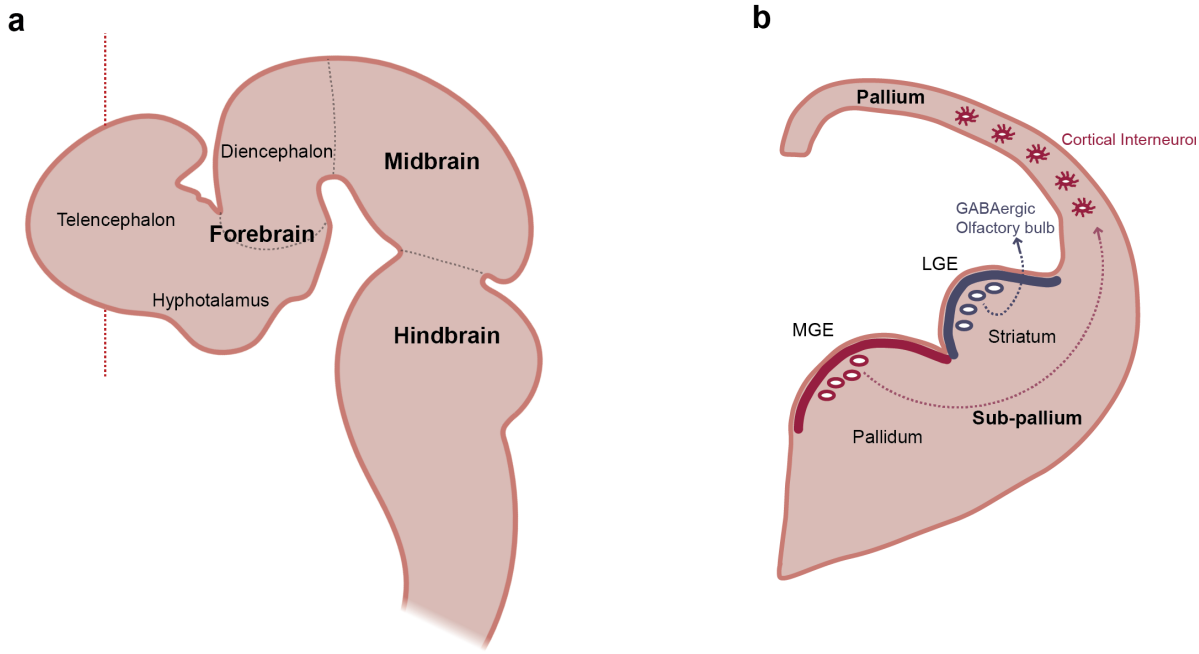


Figure 3. Schematic of the mammalian embryonic brain. a) Main brain divisions according to Puelles[15]. b) Coronal section through the telencephalon, as indicated by the dotted line in (a). The medial and lateral ganglionic eminences (MGE and LGE) are sources of GABAergic interneurons that migrate to the pallium (dorsal telencephalon) and olfactory bulbs, respectively. Furthermore, within the sub-pallium (ventral telencephalon), these eminences will give rise to the pallidum and striatum, respectively. *Illustration in (a) was created with BioRender.com*

Lastly, in most vertebrates, myelin—produced by oligodendrocytes in the central nervous system—ensheathes neuronal axons, enhancing electrical impulse transmission. The presumed absence of both myelin and oligodendrocytes in the central nervous systems of cyclostomes, suggests a gnathostome-specific innovation[42]. This view is supported by the lack of genes associated with oligodendrocyte specification and myelin synthesis in the lamprey genome. Nonetheless, reports of other myelin-associated genes and regulatory oligodendrocyte

differentiation mechanisms in lampreys raise questions about whether some regulatory toolkit for oligodendrocyte development predates the divergence of cyclostomes and gnathostomes[42,43].

1.5 The evolution of the pallium across vertebrates

The pallium, the dorsal portion of the telencephalon, has been a focal point of research because it harbors multiple structures linked with vertebrate behaviors, including cognition, learning, and memory. It is considered the most divergent part of the telencephalon, especially in amniotes. This diversity has rendered pallial homologies between species difficult to discern[27,44].

While numerous proposals have been put forth, it is conventional to categorize the embryonic pallium into four divisions: medial, dorsal, lateral, and ventral, adhering to a tetrapartite model[45,46] (Fig. 4a). The tetrapartite model, developed based on comparative embryonic genoarchitecture in amniotes, states that in mammals the medial pallium gives rise to the hippocampus and the dorsal pallium to the isocortex. The lateral and ventral pallia, which were once identified as a single pallium but have now been recognized as two distinct regions, give rise to the claustrum and insular cortex, and to specific segments of the olfactory cortex and pallial amygdala, respectively. (Fig. 4b). Modern versions of this model have included transitional architectures and molecular profiles during development and differ slightly from the original in the derivatives of each region (reviewed in Medina et al.[47]). However, the homology of these fields among amniotes remains an ongoing debate. Mapping homologies to these pallial structures is even more difficult in anamniotes. Recent single-cell transcriptomic data recovered from the amphibian telencephalon mapped neurons to four anatomical pallial regions[48]. Still, these data do not support current updates to the tetrapartite model. As all of these models were developed to explain the evolutionary relationships of pallial regions in amniotes, their application outside of this clade is challenging and controversial.

In comparative neuroanatomy, pallial homologies have also been determined by identifying the role of pallial regions in conserved neural circuits. This latter approach has provided data from anamniote brains[1,49]. However, one must consider that these "connectional fingerprints" are also subject to evolutionary change, particularly when contrasting distant species[50]. Still, a significant understanding of the divisions in each clade is derived from this methodology.

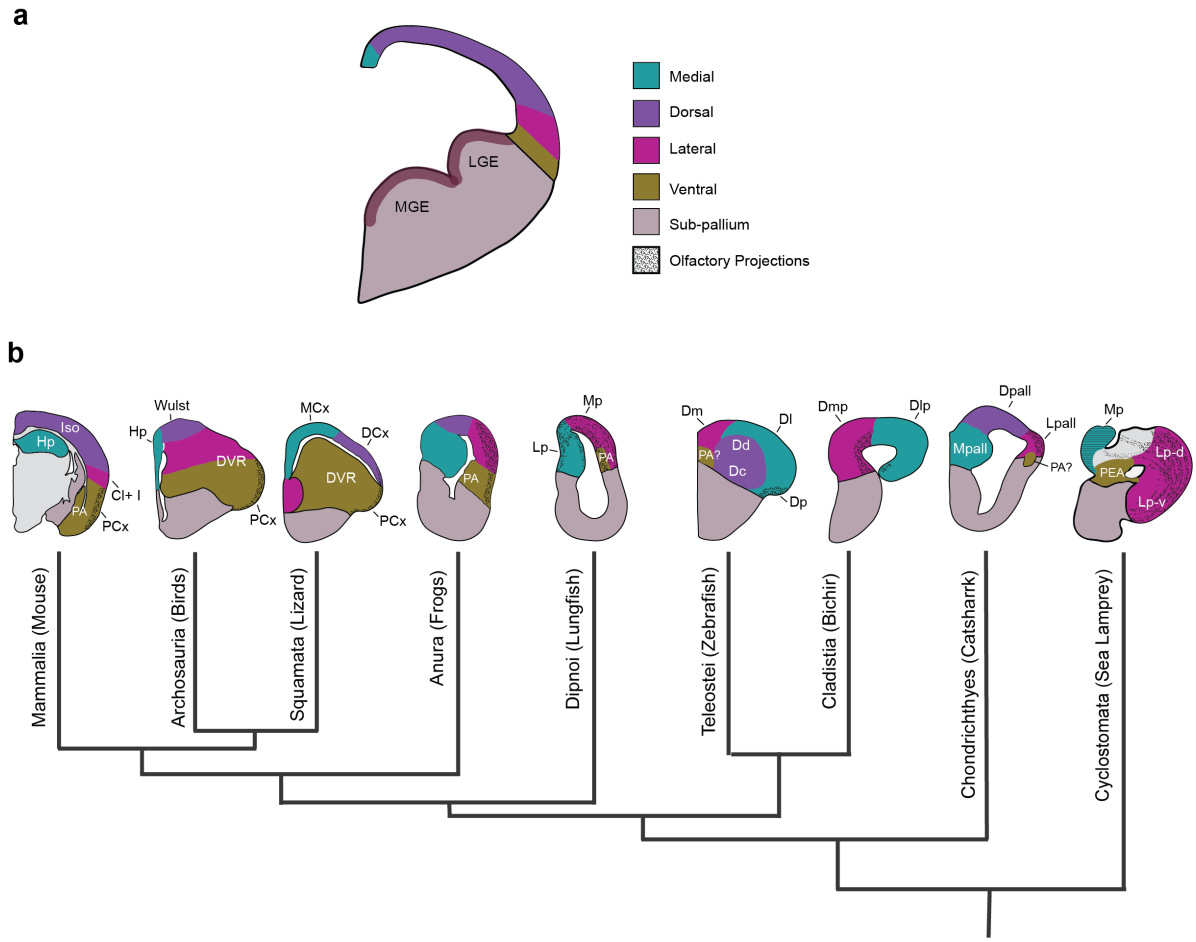


Figure 4. Vertebrate pallial divisions according to the original tetrapartite model. a) Schematic depicting the different pallial divisions in the mammalian embryo. b) Pallial divisions in the adult brains of vertebrate species. Note that multiple authors follow a tripartite model (where the ventral pallium is part of the lateral pallium) in non-amniotes. However, following the tetrapartite model, in this graphic, the pallial amygdala is identified under the ventral pallium. The scope of the olfactory bulb projections is depicted in each lineage. In lampreys, the medial pallium has been confirmed as the pre-thalamic eminence by Lamanna, Hervas Sotomayor et al[5]. **Abbreviations:** Cl: Claustrum, Dc: Area dorsalis centralis, Dcx: Dorsal cortex, Dd: Area dorsalis dorsalis, Dl: Area dorsalis lateralis, Dlp: Dorsolateral pallium, Dm: Area dorsalis medialis, Dmp: Dorsomedial pallium, Dp: Dorsoposterior pallium, Dpall: Dorsal pallium, DVR: Dorsal ventricular ridge, Hp: Hippocampus, I: Insular cortex, Iso: Isocortex, Lp-d: Lateral pallium dorsal zone, Lp-v: Lateral pallium ventral zone, LP: Lateral pallium, Lpall: Lateral pallium, MCx: Medial cortex, MP: Medial pallium, Mpall: Medial pallium, PA: Pallial amygdala, PCx: Piriform cortex (olfactory), PEA: Pallial extended amygdala. Note that some structure names are lineage-specific. Pallial divisions for different lineages follow Striedter & Northcutt[1,27], Briscoe & Ragsdale[36], Porter & Mueller[51], Tosches et al.[52], Medina & Abellán[53], Huesa et al[54]. *Mouse Illustration in (b) was created with BioRender.com*

Using the olfactory bulb's projections has been a primary approach for comparing pallial regions between species because the scope of these projections has become restricted throughout vertebrate evolution. For instance, while cyclostome brains receive olfactory bulb

inputs across the entire pallium, in other species, these inputs are limited to the lateral pallium*, leaving other areas open for different sensory inputs (Fig. 4b). Early theories about pallial evolution considered that the anamniote pallium was predominantly influenced by olfactory inputs.[7,55,56] According to these theories, the diversification of sensory inputs, with a reduced role for olfactory projections, coincided with the emergence of the dorsal pallium in amniotes. However, comparative studies of these neural projections across various vertebrates led to a reevaluation of this perspective. These studies indicated that the reduction in olfactory bulb projections is not exclusive to amniotes, challenging the previous beliefs. Consequently, comparative analyses of neural connectivity, as detailed below, suggest that the dorsal pallium may have evolved earlier than originally thought[27].

Because in cyclostomes the olfactory bulbs project nearly to the entire pallium, Northcutt and collaborators[27,55,56] proposed the presence of a lateral pallium in lampreys. This region is situated in the area where the telencephalon evaginates (Fig. 4b). Yet, this idea has been challenged because certain segments of this structure receive thalamic inputs and project to other brain areas (e.g., the midbrain and spinal cord), tentatively being identified as homologous to the mammalian dorsal pallium[57,58]. Given that the olfactory bulb's projections are restricted to relatively small parts of the pallium in sharks[27,59,60] and teleosts[61–63], and that putative dorsal pallia have been identified in these clades (Fig. 4b), a widely accepted consensus is that the dorsal pallium predates amniotes and that all vertebrates (or at least all gnathostomes) possess a tetrapartite pallium[2,64].

However, more recent comparative connectivity analyses for other vertebrates challenge the presence of a tetrapartite pallium in all vertebrates. For instance, non-teleost ray-finned fishes display diverse patterns regarding olfactory bulb projections into the pallium; most data support that in these animals, most of the pallium receives input from the olfactory bulb[61,65]. A similar pattern has been described in lungfish brains (Fig. 4b). Based on these findings and the phylogenetic positions of these lineages, an alternative hypothesis postulates that in the ancestor of vertebrates and gnathostomes, the olfactory bulb projected to most of the pallium. However, over time, this pattern became increasingly restricted in distinct lineages (sharks, teleosts, tetrapods), leading to the convergent emergence of dorsal pallia in various species[1,27].

*In amniotes, the ventral pallium receives projections from the olfactory bulb. Since the ventral and lateral pallium were previously regarded as a single region, many authors refer only to the lateral pallium in anamniotes. Because this section refers to anamniote fishes, I employ the term 'lateral pallium' following Striedter & Northcutt[1,27].

In anamniotes, other pallial divisions have attracted less attention. For instance, in lampreys, the ventral portion of the lateral pallium is seen as possibly homologous to the mammalian ventral pallium[7]. Moreover, a putative homologous of the medial pallium was identified in lampreys, based on thalamic and hypothalamic projections[56]. However, this view has been challenged by molecular data, suggesting that this structure is actually a diencephalic protrusion: the prethalamic eminence[66,67]. With limited molecular evidence, many researchers lean towards the connectivity-based interpretation. In other anamniotes, molecular data has unveiled signs of these divisions (medial and ventral)* across vertebrates, or at the very least, the genetic toolkit necessary for their formation[1,2]. Yet, for many lineages, such data remains sparse or non-existent, often relating to only a handful of genes, restricting the conclusions that can be drawn from them. A comprehensive approach, encompassing systematic comparisons across a wide spectrum of vertebrate lineages with a strong representation of anamniotes, coupled with extensive genomic data, is essential to genuinely understand the origins and evolution of the pallium.

1.6 Comparative neuroscience and the evolution of cell types

A core challenge in comparative neuroscience lies in the approach employed: given the vast complexity of the brain, the methods and criteria used to compare it across species can significantly impact the outcomes and conclusions drawn[1]. While brain structures may appear homologous at a superficial level, a deeper examination, as exemplified by the pallium[27], complicates our ability to distinguish between conservation, convergence (or parallelism), and innovation.

In recent years, the inclusion of strategies prioritizing cell composition comparisons has yielded valuable insights into the evolution of structures[1,68–70]. Considering that brain tissues are typically heterogeneous in their cell composition and that many of the debates about brain homology consist of single cell types, delving into cell evolution might contribute to elucidate existing controversies in neuroscience. However, this approach is not without complications. Cells, like other biological units, evolve and are influenced by selective pressures. Historically, cell identity has been determined based on morphology and function, mostly on a handful of model organisms. While these classifications have strengthened multiple research endeavors,

*The original tetrapartite model categorizes the pallial amygdala as a derivative of the ventral pallium. However, in anamniotes, it is frequently not referred to as such in the scientific literature.

their application in comparative biology is limited. Notably, they often fail to distinguish between homologous and convergent cell identities[70–72].

To accurately identify and compare cells across species, we need a standardized evolutionary approach to cell identification, ideally nested within a phylogenetic framework. The advent of advanced technologies, such as single-cell RNA sequencing, has strengthened a concept of cell type that accommodates both cell evolution and identity[71,73–75]. This evolutionary concept centers on understanding the processes underlying the origin and evolution of cell types across species. It defines a cell type as "a group of cells in an organism that evolves collectively, partially independent of other cells, and is evolutionarily more related to each other than to different cells"[71].

This perspective introduces the idea of Core Regulatory Complexes (CoRCs). These are conglomerates of transcription factors (termed 'terminal selectors') that enable and sustain a cell's unique gene expression program. By regulating the expression of effector genes, these selectors essentially dictate cell identity. Importantly, this concept also accounts for the hierarchical clustering generated by transcriptomic similarities, paving the way for the recognition of broader classification categories, such as cell type families[71,76]. This has significant implications for understanding cell evolution across phylogenetic lineages. Nonetheless, applying this concept presents numerous challenges, including comparing transcriptomic data from diverse cells across evolutionarily distant species and reconciling cell evolution with the broader evolution of organs. Recent advancements have successfully addressed the first limitation by developing methods that go beyond solely considering one-to-one orthologues[77–79], an approach commonly used in comparative genomics but constrained when the species being compared are not closely related or when significant genomic changes have occurred since their divergence. The other challenge, integrating cell and organ evolution, has increasingly gathered attention, particularly as current research endeavors shed light on it[80–82]; however, comprehensive conceptual frameworks are not yet completely established.

2 Aim of the research

This thesis aims to address long-debated controversies in comparative neuroscience, by providing a comprehensive analysis of the cellular origin and molecular evolution of the vertebrate brain. The current work is the result of two intertwined projects, where one paved the way for the other. Given their aligned and complementary scientific goals, I have merged them into a single and inclusive narrative, with the specific following aims:

- To infer the ancestral cellular landscape of the vertebrate brain through a comparative analysis of cell atlases from the lamprey and mouse.
- To reconstruct the molecular core of the vertebrate brain by identifying conserved transcription factors profiles across anamniote species including lamprey, catshark, spotted gar and lungfish.
- To identify homologous cell classes across anamniote species and their molecular programs.
- To explore the origin and diversification of cells and tissues, specifically focusing on oligodendrocytes, and inhibitory and excitatory neurons within the telencephalon.

3 Results

The following sections describe the work carried out in collaboration with other members of the Kaessmann Lab. For a detailed description of the individual duties, please refer to the methods section. In brief, all larval and adult lamprey analyses were conducted by Dr. Francesco Lamanna, using data generated by Dr. Phil Oel and myself. Additionally, I performed all the work involving spatial transcriptomics and microscopy. The main findings of this collaboration are publicly available in a publication where I am co-first author[5]. Independently, I generated and analyzed all lamprey embryonic data presented in this thesis.

The data production, quality control pipelines, and broad cell annotation for the catshark, spotted gar, and lungfish were undertaken by Matthias Janeschik, Dr. Kerry Lynn Gendreau, and myself, respectively. However, all exploratory analyses and plots depicted below for these species were generated by me. Therefore, through the results and discussion section when I describe the work carried out for adult and larval lamprey, I used the plural first-person pronoun. For all other results (lamprey embryos, catshark, spotted gar and lungfish) I used the singular first-person pronoun.

3.1 Cellular diversity in the lamprey brain

The primary goal of the first part of this research was to gain a comprehensive overview of cell type diversity in the lamprey brain. By drawing comparisons with other vertebrates, we also aimed to elucidate the underlying cell type diversity of the vertebrate brain and to uncover its ancestral molecular core. To achieve this, we began by constructing cell type atlases of the larval and adult lamprey brains at single-cell resolution, encompassing 72,810 cells for adults and 86,571 for larvae (Fig. 5a-d). Despite differences in cluster structure and organization, our findings indicate that the larval and adult atlases are comparable in terms of cell diversity and molecular patterning. As evidenced by cell annotation and localization (Fig. 5a-b,e-h), both non-neuronal and neuronal broad cell classes are present at both stages. Adults, however, are considered to better represent ancestral states for all vertebrates, as the larval stage is derived in the lamprey lineage[83].Therefore, unless specified otherwise, all results described below correspond to the adult dataset but are applicable to the larval stage as well.

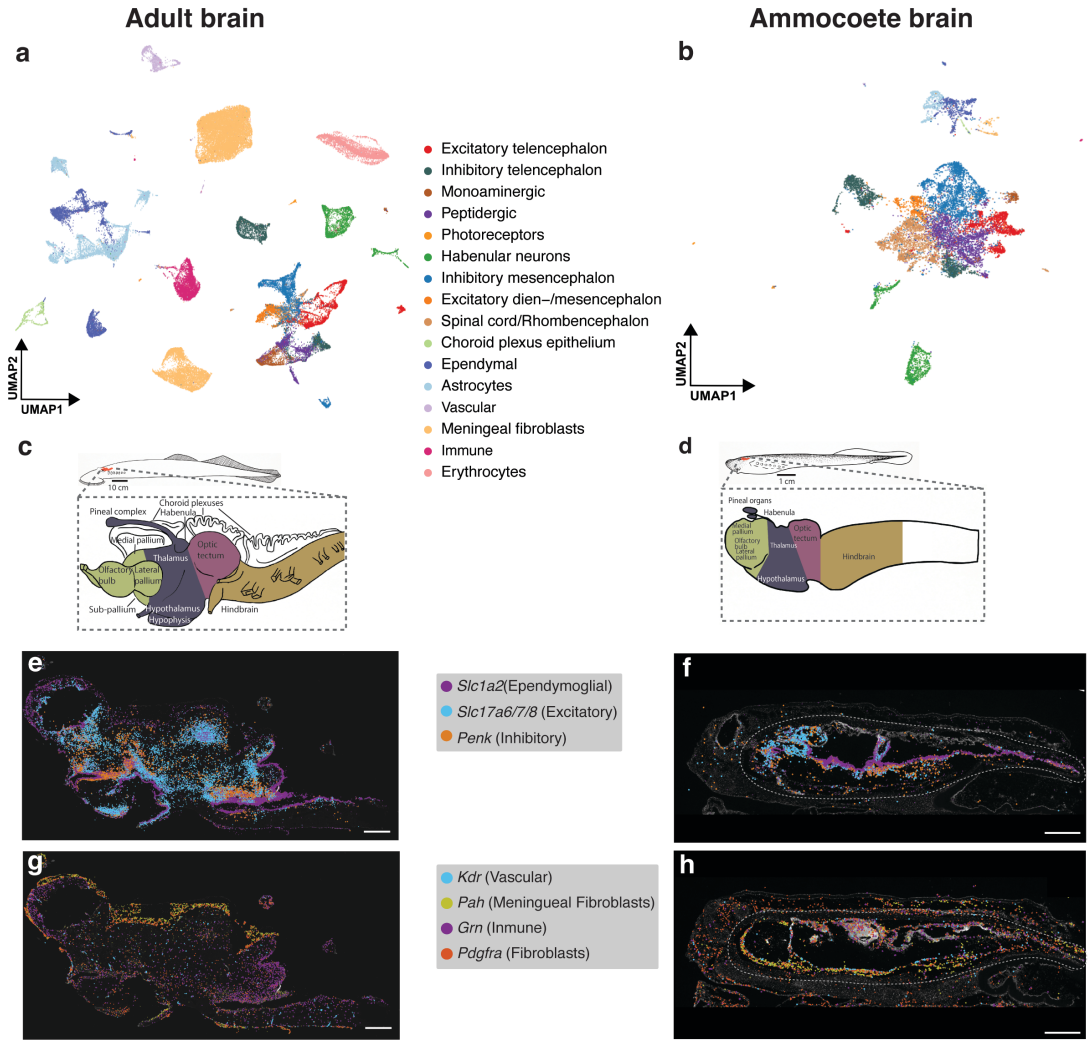


Figure 5. Overview of the lamprey brain atlas. a-b) UMAP representation of adult (a) and larval (b) cells (all single-cell RNA sequencing data combined) colored according to their respective cell type classes. c-d) Schematics of the adult and larval brains, respectively. Colors indicate the different regions dissected for this study. e-h) Sagittal sections (oriented as in b and c) of both adult and larval brains, displaying ISS (*in situ* sequencing) maps of genes marking various cell type groups. As the larval sections correspond to the heads of the animals, the brain is outlined with a dashed line. Scale bars represent 500 µm. *Figures (a-d) have been adapted from Lamanna, Hervas-Sotomayor et al*[5]*, license under Creative Commons CC BY.*

Cell type relationships based on gene expression distances tend to unveil a hierarchical organization of cells. This implies that individual cell populations group to form distinct cell types, which then assemble into families. These families further aggregate to constitute classes, and so on. The lamprey brain demonstrates this hierarchical organization (Fig. 6a), where similar cells group into cell types, and these, in turn, form families and classes. This hierarchy eventually culminates in the differentiation of cell diversity into neuronal and non-neuronal classes, akin to what is observed in other species[84,85], indicating that this dichotomy is deeply conserved.

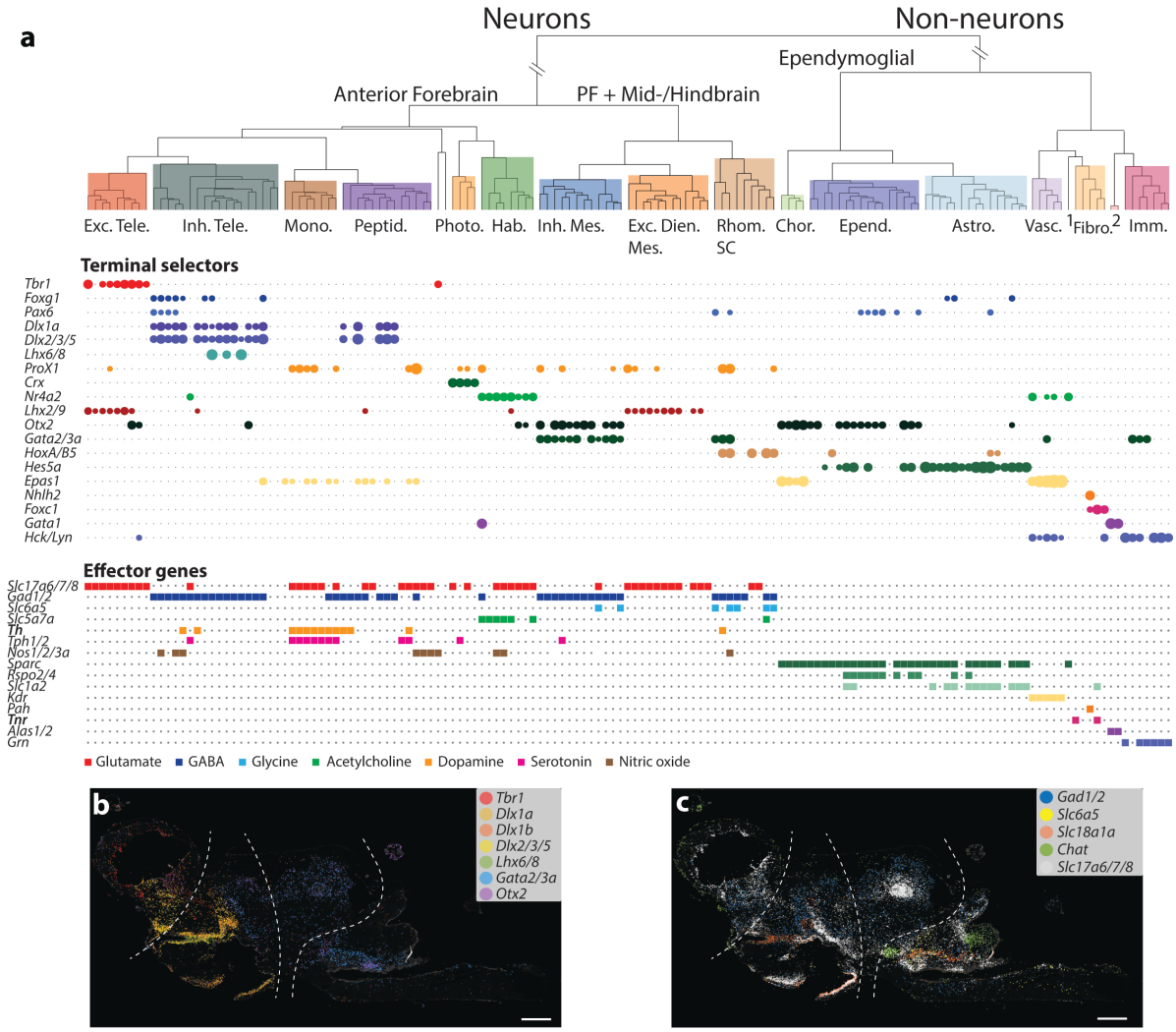


Figure 6. Lamprey adult brain cell classes and molecular specification. a) At the top, a dendrogram illustrating the relationships between the identified cell type classes, which are highlighted by colored boxes. At the middle, the expression of transcription factors (terminal selectors) within each cell type class is shown; the sizes of the circles are proportional to the number of cells expressing each gene. At the bottom, a binary representation of effector genes expression (presence/absence) is provided, based on whether a gene is differentially expressed within the corresponding cell type class. b-c) Sagittal sections (same orientation as Fig.5c) of the adult brain display ISS maps of terminal selectors (b) and effector genes (c). Dashed lines demarcate the four main brain regions as illustrated in Fig.5c. Abbreviations: PF, posterior forebrain; SC, spinal cord; 1, PNS glia; 2, erythrocytes. Scale bars represent 500 µm. Note that because the majority of species studied in this thesis are gnathostomes, gene names are not capitalized as done for lamprey genes in Lamanna, Hervas-Sotomayor et al[5]. Instead gene names are presented according to their mouse orthologues for consistency, following the format used by Woych et al.[48]. *Adapted from Lamanna et al[5], license under Creative Commons CC BY.*

A significant challenge was encountered when attempting to discern the boundaries of these hierarchical units, mainly because the demarcations between them are not always clear, particularly when examining cell populations or types. Such ambiguities are not random but

rather an inherent property of the evolutionary process, because cells as other evolutionary units tend to become more distinct as they diverge. Consequently, to accurately classify cell types and decipher this hierarchy, we adopted an approach that involved compiling as much evidence of cellular identity as possible. By applying the concept of CoRs[71] and using the spatial localization of cells, we effectively annotated cell hierarchical relationships. CoRs, which consist of specific terminal selectors or transcription factors with concerted expression profiles, are instrumental in defining cell identity across all hierarchical levels[71]. After identifying these gene sets, spatial information confirmed their co-expression while facilitating the localization and annotation of cells (Fig. 6).

Our analysis has illuminated the role of CoRs in determining cell identity across various hierarchical levels in the lamprey brain. We have identified sets of transcription factors and effector genes with co-expression patterns unique to cell types, families and classes (Fig. 6a). Although the causality of these co-expression profiles was not experimentally validated, the spatial expression patterns of these genes (Figs. 6b-c) suggest that they may constitute, at least in part, the identity-determining CoRs.

3.2 Homologous cell type classes and families across vertebrates

The brain cell diversity of the catshark, gar, and lungfish comprises the same broad cell classes described for other vertebrate species (Suppl. Fig. 1), encompassing both neuronal and non-neuronal classes. As with lampreys, cell class identity in these species is indicated by the co-expression of gene sets, which likely determine cell identity at multiple hierarchical levels. It is important to note that for these three species, I used data generated only from the telencephalon (37,620 cells for catshark; 24,890 cells for spotted gar; and 20,082 cells for lungfish), whereas the lamprey dataset encompasses the entire brain. Owing to this difference and the ongoing nature of data generation for these species, I have conducted only a preliminary exploration, identifying genes differentially expressed in specific cell classes. Similarly, the interspecies comparison discussed below was restricted to broad cell classes for the same reasons. Nevertheless, the results suggest that, mirroring the findings in lamprey and mouse cells, the identities of broad cell classes in sharks, gars, and lungfish are also characterized by the expression of particular gene sets.

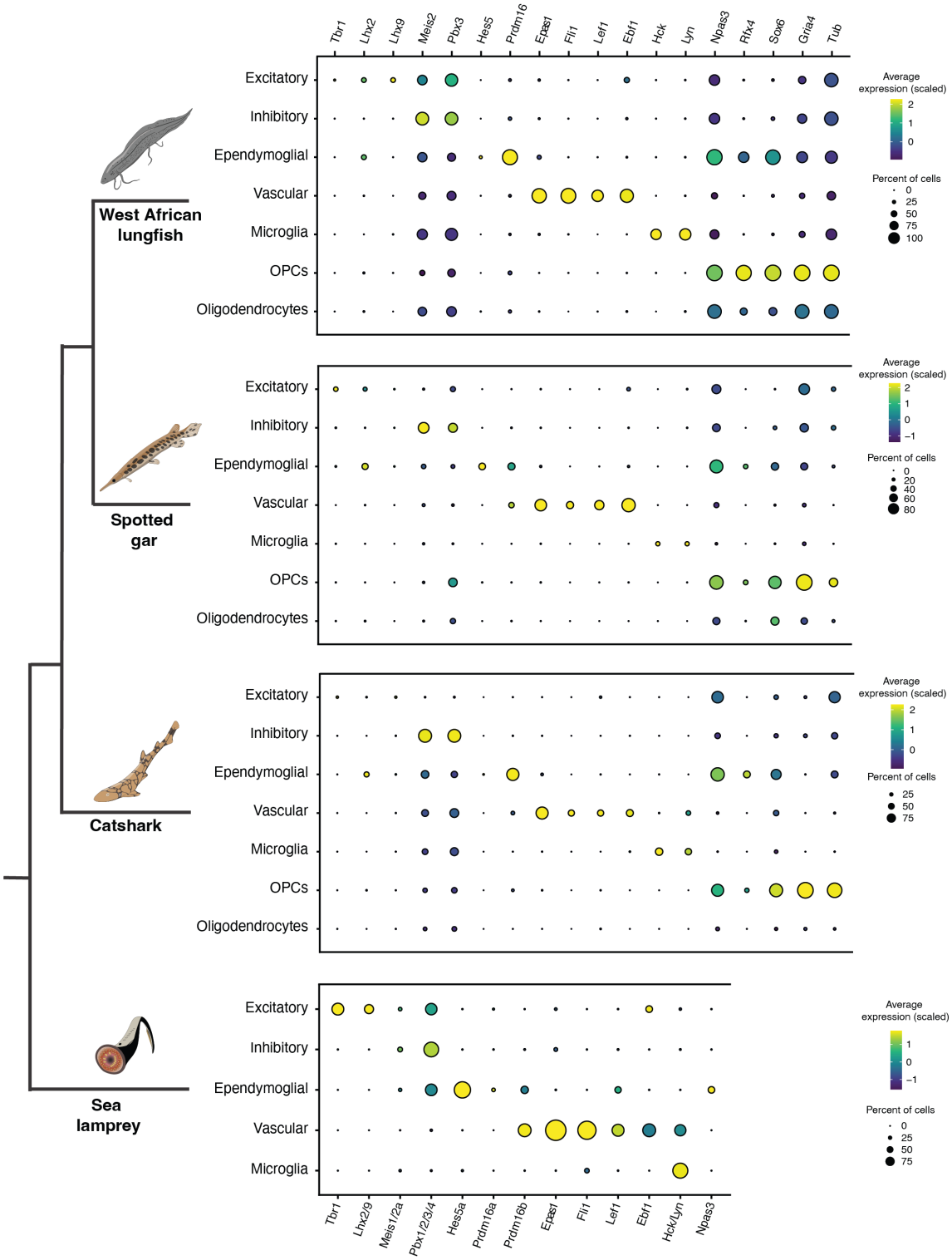


Figure 7. Conserved expression of transcription factors across vertebrates. Dot plots illustrate the expression of transcription factors within the same broad neural cell classes in lungfish, spotted gar, cat shark, and sea lamprey. Since lampreys lack oligodendrocytes, gene sets typically associated with these cells are not depicted for this species. In instances where genes have duplicated in the common ancestor of gnathostomes and ancestral expression has been conserved, the expression of both paralogues is indicated. Fibroblasts were excluded from this figure because they were not recovered in the spotted gar.

The expression patterns of some gene orthologues is remarkably conserved throughout the vertebrate clade, as suggested by the number of transcription factors and effector genes that show concordant expression profiles across the four species (Fig. 7-8). These findings, particularly the conserved expression profiles of transcription factors (Fig. 7), suggest that these broad cell classes are homologous and that, at least part of the CoRs that define the class identity, are conserved across vertebrates, likely representing an ancestral molecular blueprint. However, it is important to note that many genes identified as differentially expressed between cell classes do not exhibit a strongly conserved expression pattern across species (Suppl. Fig. 1). This outcome is unsurprising given that cell classes have been evolving independently within each clade for tens of millions of years. Consequently, it is expected that they would accumulate significant differences in the expression patterns of numerous genes (particularly effector genes) over such a timescales (Fig. 1). The challenge lies in discerning which of these genes are part of an ancestral molecular core, useful for homologizing cells, and which ones encode clade-specific identities, shedding light on the evolutionary trajectory of these cells after the divergence of clades. Furthermore, this observation warrants a note of caution: traditional marker genes used for specific cell classes might not necessarily indicate the same cell class in a different species. Alternatively, such markers might identify an entirely different cell. Relying on a limited set of markers for cell identification could inadvertently lead to incorrect homologations. Therefore, adopting a more systematic and comprehensive approach, in addition to the use of markers, is crucial for accurately identifying homologous cells.

An illustration of the potential pitfalls associated with relying on marker genes is the expression of *Slc17a7* (*Vglut1*) and *Slc17a6* (*Vglut2*) (Fig. 8, red arrows). In lampreys, only one copy of this gene exists while in gnathostomes, it underwent a duplication event. Both paralogous genes encode for the main glutamate transporters, traditionally considered markers for glutamatergic cells. In mammals, their expression seems to be complementary with only minimal overlapping[84,86]. In these animals, *Slc17a7* is primarily expressed in the pallium, including the isocortex, hippocampus, amygdala, and cerebellum, while *Slc17a6* is found in specific cortical layers, as well as subcortical regions, the thalamus, and brainstem[86]. Therefore traditionally, *Slc17a7* has been used to identify glutamatergic cells in the telencephalon. In lampreys, the single copy of this gene (*Slc17a6/7*) shows generalized glutamatergic expression. This is mirrored in the shark's telencephalon, where both genes mark excitatory neurons (Fig. 8, red arrows). However, in the gar telencephalon, all glutamatergic neurons express *Slc17a6*, while *Slc17a7* expression appears absent. Conversely, in the lungfish, *Slc17a7* is the general

glutamatergic marker, with *Slc17a6* expressed in only a few cells. This pattern is similar in the salamander telencephalon, where *Slc17a6* is exclusively expressed in the glutamatergic cells of the amygdala[48,82]. Remarkably, in turtles, *Slc17a6* and *Slc17a7* are coexpressed in all pallial glutamatergic neurons[52]. These results suggest that the regulation of these transporters has shifted multiple times during vertebrate evolution, while maintaining their ancestral role in excitatory neurons.

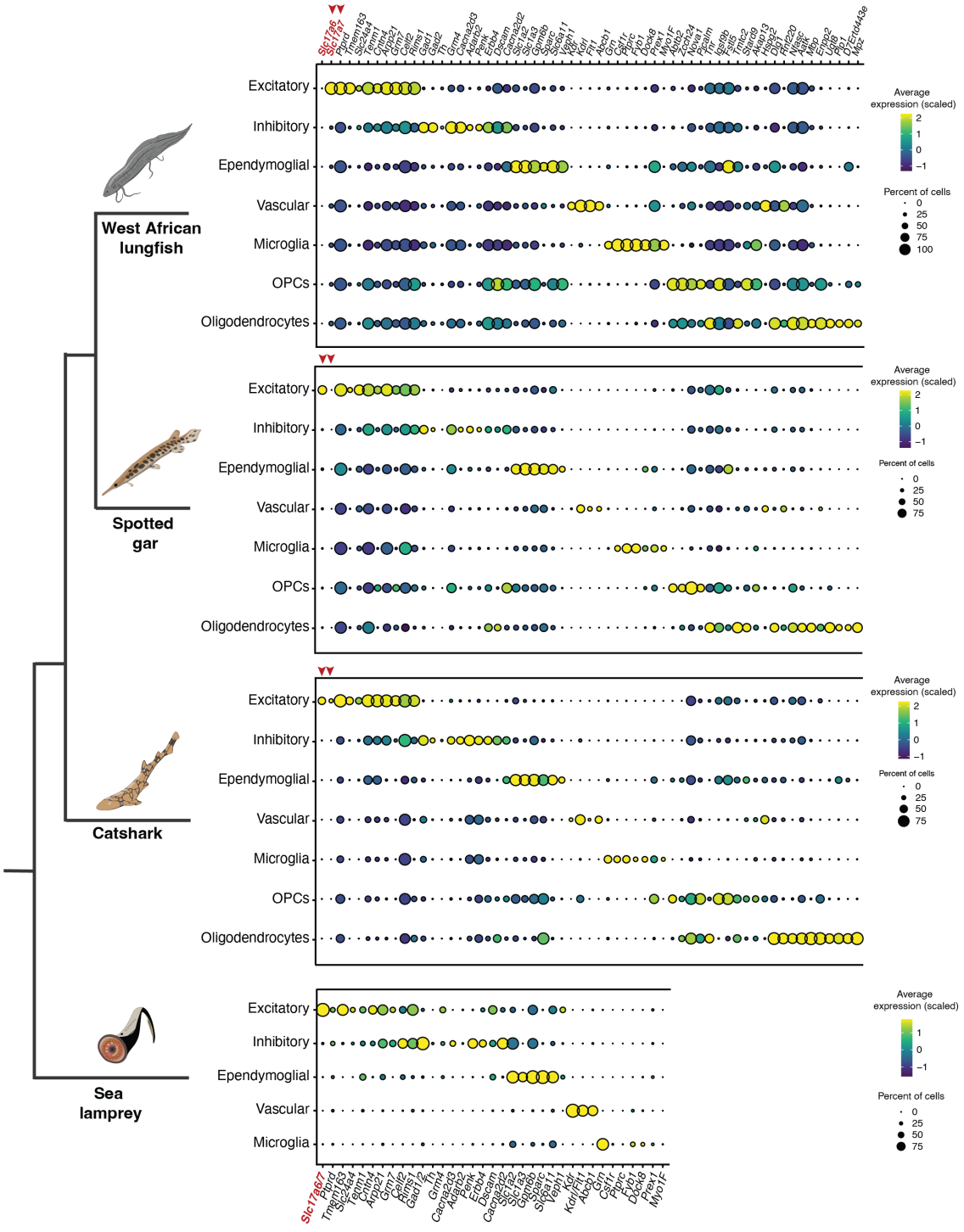


Figure 8. Conserved expression of effector genes across vertebrates. Dot plots display the expression of effector genes across similar broad neural cell classes in lungfish, spotted gar, cat shark, and sea lamprey. Since lampreys lack oligodendrocytes, gene sets typically associated with these cells are not depicted for this species. In instances where genes have duplicated in the common ancestor of gnathostomes and ancestral expression has been conserved, the expression of both paralogues is indicated. Fibroblasts were excluded from this figure because they were not recovered in the spotted gar. Red arrows highlight the variable expression of *Slc17a6* and *Slc17a7* in glutamatergic neurons amongst the species studied.

Apart from identifying cell-specific gene sets, we performed systematic comparisons between lamprey and mouse datasets (Fig. 9a-f), using various approaches ("Meta-gene"[79] and "SAMap"[77]) that are not limited to only one-to-one orthologues, a considerable limitation when comparing phylogenetically distant species (see Methods). For these analyses, we used finer resolution clusters to evaluate the further conservation of cell groups at a higher hierarchical level. I conducted similar systematic comparisons for other species as well, but only for excitatory neurons from the telencephalon (discussed in section 3.4.2.2). The conservation of gene modules across species is also reflected in the results from these systematic comparisons. We found robust correlations between the cell type groups from lamprey and mouse brains; this is true for both glia and neurons (Fig. 9a-b). This observation was further supported when we co-embedded the atlases of the two species and identified similarity values between these groups (Fig. 9c-f). Based on these correlations and similarity values, we propose that these cell groups represent vertebrate cell type families that were likely present in common vertebrate ancestors. Notably, however, mouse cerebellar neurons show no correlation with any neurons from the lamprey brain, as indicated by the lack of support for such a correlation in the dendrogram (Fig. 9b). This conclusion is further supported by the absence of lamprey cell clusters that specifically express markers associated with Purkinje or granule cells of the cerebellum (Fig. 5-6). These findings confirm the absence of cerebellar nuclei in the lamprey brain[87,88].

While delving deeper into the cell hierarchy, challenges arose when comparing directly cell types. At an even finer resolution, correlations and embeddings become more complex and difficult to interpret, as one-to-one cell type relationships have only rarely been observed between distant species. Nevertheless, we discerned few one-to-one cell cell type relationships between lamprey and mouse. For instance, hypendymal cells of the sub-commissural organ (SCO), which express SCO-spondin (SSPO, Fig. 9g-h).

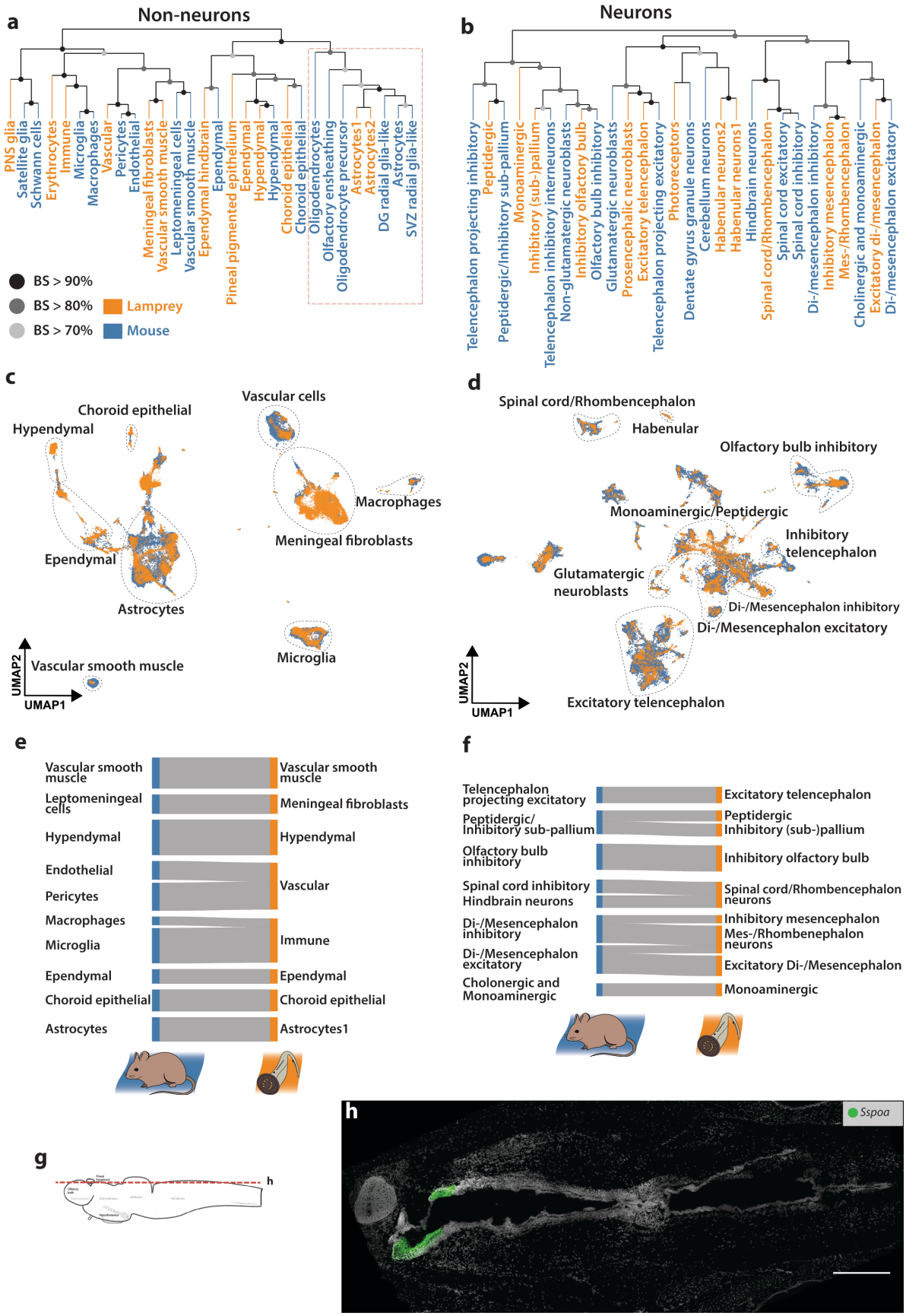

a
Non-neurons
PNS glia
Satellite glia
Schwann cells
Erythrocytes
Immune
Microglia
Macrophages
Vascular
Pericytes
Endothelial
Meningeal fibroblasts
Vascular smooth muscle
Leptomeningeal cells
Vascular smooth muscle
Ependymal hindbrain
Ependymal
Pineal pigmented epithelium
Ependymal
Hypendymal
Hypendymal
Choroid epithelial
Choroid epithelial
Oligodendrocytes
Olfactory ensheathing
Oligodendrocyte precursor
Astrocytes1
Astrocytes2
DG radial glia-like
Astrocytes
SVZ radial glia-like
BS > 90%
BS > 80%
BS > 70%
Lamprey
Mouse
b
Neurons
Telencephalon projecting inhibitory
Peptidergic
Peptidergic/Inhibitory sub-pallium
Monoaminergic
Inhibitory (sub-)pallium
Telencephalon inhibitory interneurons
Non-glutamatergic neuroblasts
Inhibitory olfactory bulb
Olfactory bulb inhibitory
Glutamatergic neuroblasts
Prosencephalic neuroblasts
Excitatory telencephalon
Telencephalon projecting excitatory
Photoreceptors
Dentate gyrus granule neurons
Cerebellum neurons
Habenular neurons2
Habenular neurons1
Hindbrain neurons
Spinal cord/Rhombencephalon
Spinal cord excitatory
Spinal cord inhibitory
Di-/mesencephalon inhibitory
Inhibitory mesencephalon
Mes-/Rhombencephalon
Cholinergic and monoaminergic
Excitatory di-/mesencephalon
Di-/mesencephalon excitatory
c
Choroid epithelial
Hypendymal
Vascular cells
Macrophages
Meningeal fibroblasts
Ependymal
Astrocytes
Microglia
Vascular smooth muscle
UMAP2
UMAP1
d
Spinal cord/Rhombencephalon
Habenular
Olfactory bulb inhibitory
Monoaminergic/Peptidergic
Inhibitory telencephalon
Glutamatergic neuroblasts
Di-/Mesencephalon inhibitory
Di-/Mesencephalon excitatory
Excitatory telencephalon
UMAP2
UMAP1
e
Vascular smooth muscle
Leptomeningeal cells
Hypendymal
Endothelial
Pericytes
Macrophages
Microglia
Ependymal
Choroid epithelial
Astrocytes
Vascular smooth muscle
Meningeal fibroblasts
Hypendymal
Vascular
Immune
Ependymal
Choroid epithelial
Astrocytes1
f
Telencephalon projecting excitatory
Peptidergic/ Inhibitory sub-pallium
Olfactory bulb inhibitory
Spinal cord inhibitory
Hindbrain neurons
Di-/Mesencephalon inhibitory
Di-/Mesencephalon excitatory
Cholonergic and Monoaminergic
Excitatory telencephalon
Peptidergic
Inhibitory (sub-)pallium
Inhibitory olfactory bulb
Spinal cord/Rhombencephalon neurons
Inhibitory mesencephalon
Mes-/Rhombenephalon neurons
Excitatory Di-/Mesencephalon
Monoaminergic
g
h
h
Sspoa

Figure 9. Comparisons between lamprey and mouse brain atlases. a-b) Dendrograms showing the gene expression distance (Pearson's r) for transcription factor genes of non-neuronal (a) neuronal (b) cell type groups from both species. Bootstrap support is indicated (n = 1,000). c-d) SAMap results displaying UMAPs of embedded non-neuronal (c) and neuronal (d) cells from both species. Erythrocytes and oligodendrocytes have been excluded from the lamprey and mouse datasets, respectively. e-f) Sankey diagrams demonstrate the relationships between non-neuronal (e) and neuronal (f) cell type groups across the two species, based on SAMap mapping scores (minimum = 0.1; maximum = 0.65). The width of the links corresponds to the mapping scores. DG represents the dentate gyrus and SVZ represents the sub-ventricular zone. g) Schematic of the brain illustrates the plane of section shown in h. h) Horizontal section through the dorsal part of the larval brain. Scale bars represent 500 µm. *Adapted from Lamanna, Hervas-Sotomayor et al[5], license under Creative Commons CC BY.*

3.3 Origin and molecular evolution of myelinating glia

Our cross-species comparisons yielded insights into conserved elements across vertebrates and the origins of structures, cells, and molecular pathways unique to this clade. One significant finding pertains to the origin of myelinating glia, which have been reported exclusively in gnathostomes. In the central nervous system of these animals, oligodendrocytes produce myelin—an insulating substance crucial for fast and efficient nervous impulse transmission[2]. Despite the absence of oligodendrocytes in cyclostomes, leading to unmyelinated axons, genes linked to myelin synthesis have been identified in the genomes of both lamprey[42,89] and hagfish[42,90].

The lamprey-mouse correlations for non-neuronal cells revealed a surprising relationship among glial cells (Fig. 9a, dotted square). Although glial cells from both species form a supported cluster (bootstrap > 80) within the dendrogram, mouse oligodendrocyte progenitor cells (OPCs) exhibit a higher correlation with both mouse and lamprey astrocytes than with mouse oligodendrocytes. This correlation is further supported by the presence of OPC and oligodendrocyte-specific gene expression in lamprey astrocytes (Fig. 10). Previously, oligodendrocyte-associate molecular elements have been identified during lamprey gliogenesis[42] and astrocytes have been reported to ensheath the axons of neurons in this species[91]. These findings collectively suggest that modern oligodendrocytes likely evolved from shared ancestral glial cells akin to present-day astrocytes, possibly through a process of divergence and specification.

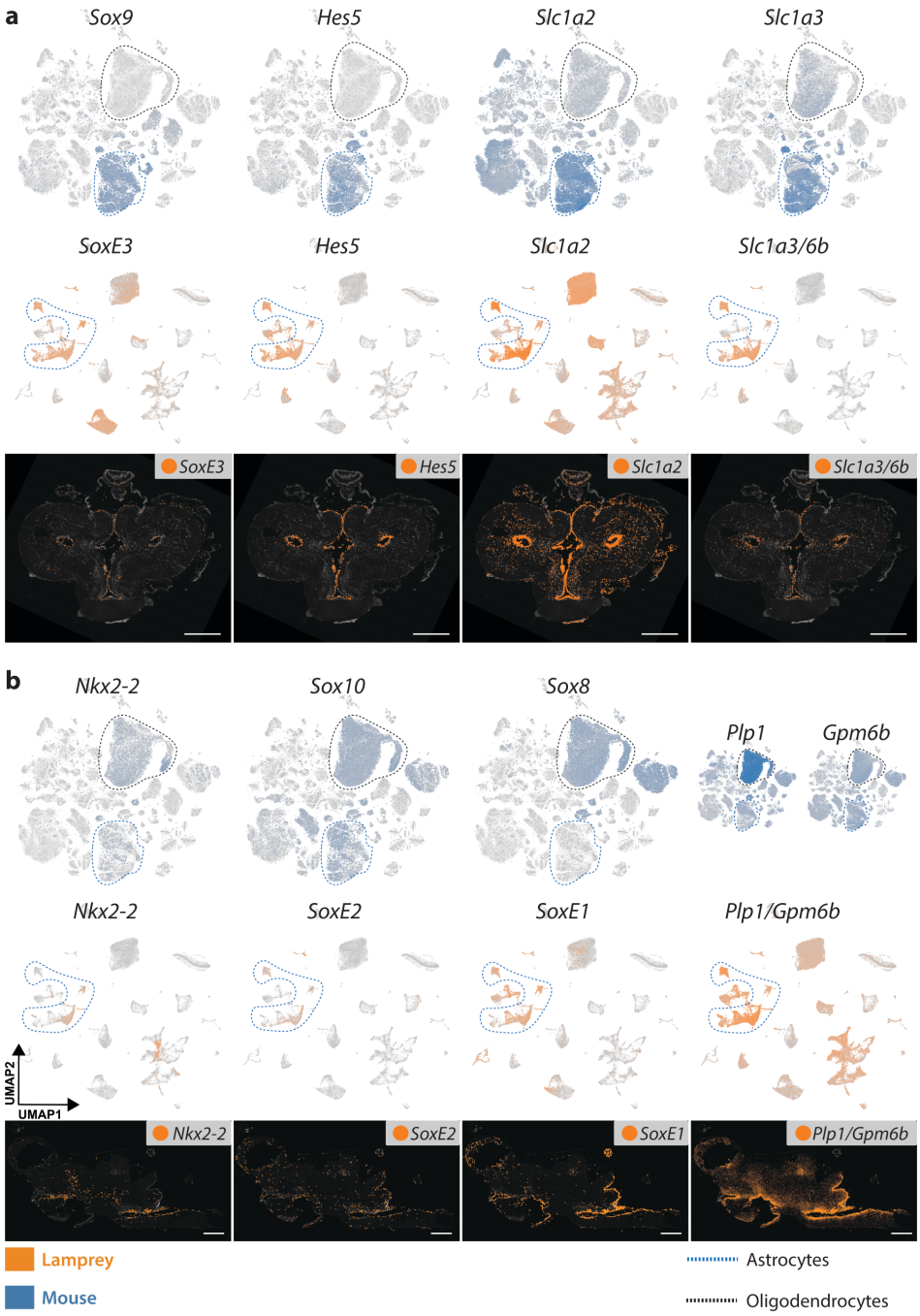


Figure 10. Expression of astrocyte-specific and oligodendrocyte-specific genes. a-b) UMAPs depicting the expression of genes specific to astrocytes (a) and oligodendrocytes (b) with orthologous genes in the mouse (top panels) and lamprey (middle panels) brain atlases. At the bottom, panels showing ISS maps of the adult lamprey brain for the corresponding genes, in coronal sections of the telencephalon (a) and sagittal sections of the entire brain (b). Refer to Fig. 14c,f and Fig. 5b. for the schematic of sections shown in a and b, respectively. Scale bars represent 500 µm. *Adapted from Lamanna, Hervas-Sotomayor et al*[5]*, license under Creative Commons CC BY.*

To further explore the evolution of present-day glia across vertebrates and the accompanying genomic pathways, I conducted a comparative analysis including data from the catshark,

spotted gar, and lungfish astrocytes and oligodendrocytes (Fig. 11). This exploration identified numerous genes, including transcription factors and effector genes, with conserved expression in this glial class across vertebrates. These findings indicate a shared gene core in this vertebrate glia, likely present in the ancestral cell type that gave rise to current oligodendrocytes and astrocytes. Additionally, this exploration provided important insights into how the emergence of genes in the gnathostome lineage influenced the evolution of oligodendrocytes.

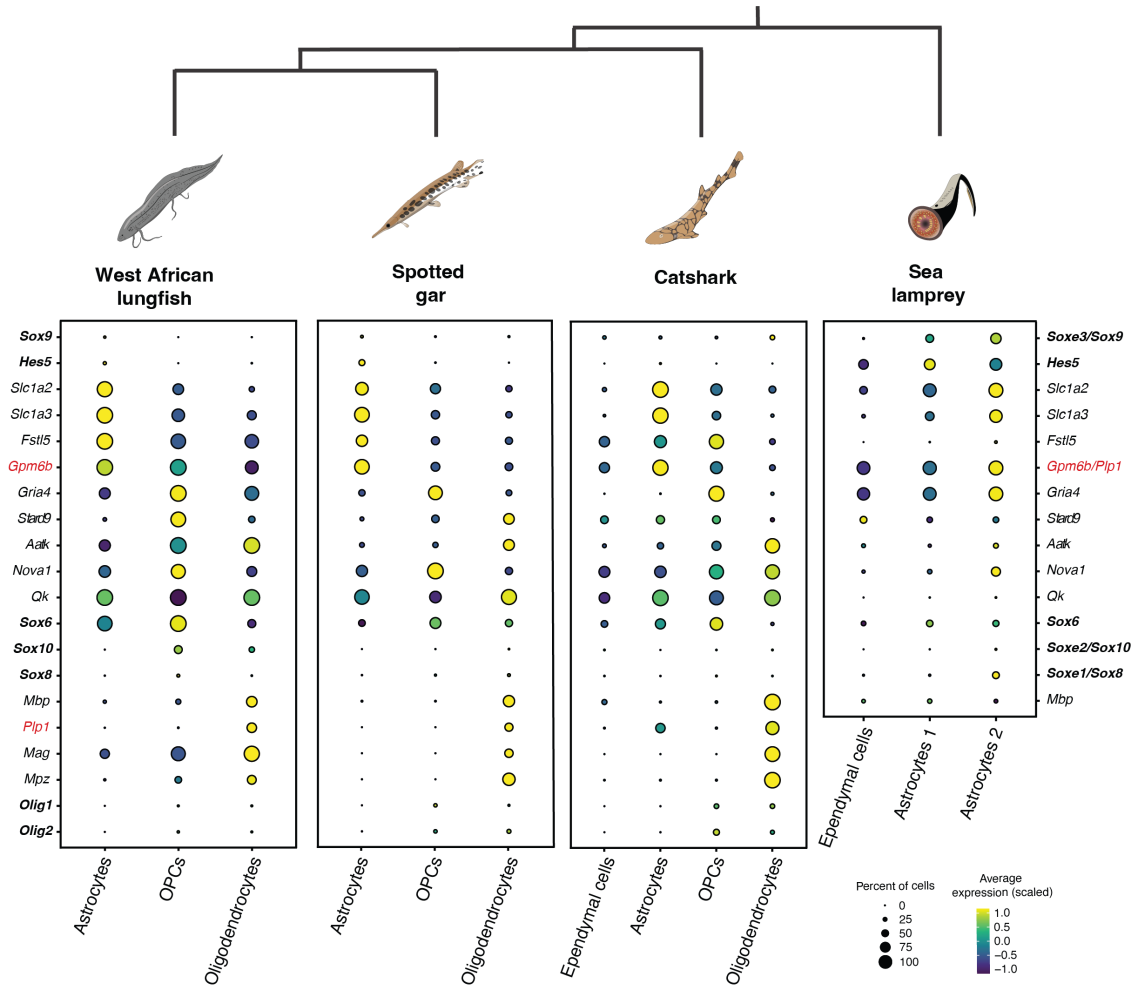


Figure 11. Expression of astrocyte-specific and oligodendrocyte-specific genes across vertebrates. This figure presents only those genes with a conserved expression profile across gnathostomes. It includes all genes depicted in Fig.10 with the exception of *Nkx2-2,* which is not expressed in lungfish cells. Transcription factors are highlighted in bold. The paralogous genes *Gpm6b* and *Plp1* are distinguished in red.

To gain a more comprehensive understanding of how gene expression has changed throughout the evolution of this vertebrate glial class, I classified the genes depicted in Figures 10 and 11 into four groups, according to their expression level (Fig. 12). This categorization is based on an assessment of expression levels derived from the exploratory plots presented in Figure 11.

Note that the expression of individual genes from each group is not identical across species; instead, the concerted expression of these genes appears to be shared among different species.

- Group A: Highly expressed by astrocytes across all species.
- Group B: Highly expressed by lamprey astrocytes and gnathostome astrocytes and OPCs/oligodendrocytes.
- Group C: Highly expressed by lamprey astrocytes and gnathostome OPCs/oligodendrocytes but lowly expressed by gnathostome astrocytes.
- Group D: Expressed primarily by OPCs/oligodendrocytes in gnathostomes. Many of these genes are absent in the lamprey genome, indicating their likely critical role in the specification of OPCs/oligodendrocyte identity.

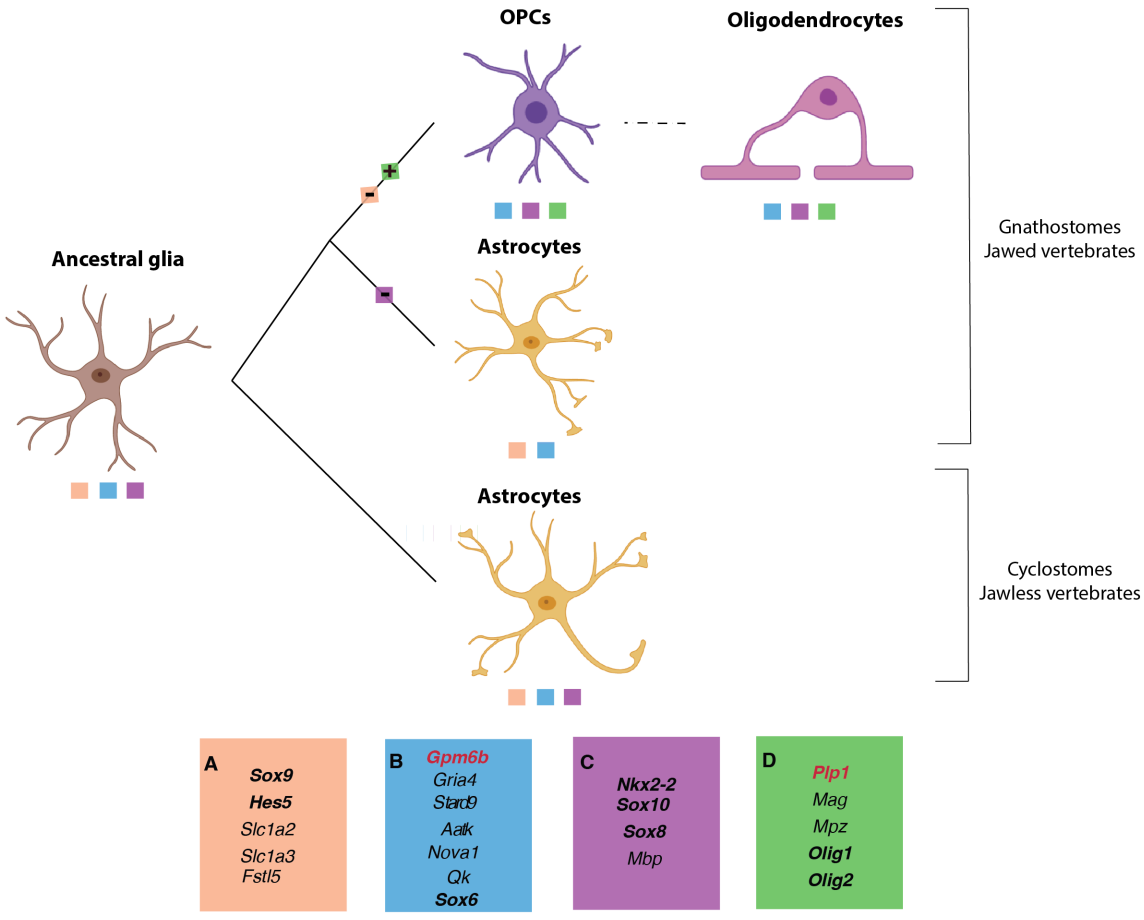


Figure 12. Hypothetical model of oligodendrocyte evolution. Genes with concerted expression are grouped within colored boxes. Below each illustration, boxes indicate gene groups expressed by extant cell types. An addition or significant increase in the expression of a gene group is denoted by a plus sign (+), while a significant decrease is represented by a minus sign (-). Transcription factors are highlighted in bold. The paralogues *Gpm6b* and *Plp1* are represented in red. Note that only one

copy of this gene is present in extant lamprey astrocytes and likely in the ancestral glial (represented by *Gpm6b* in box B). Abbreviations: OPCs: oligodendrocyte precursor cells. C*reated with BioRender.com.*

I synthesized all observations mentioned above into a hypothetical model that describes the molecular evolution of astrocytes and oligodendrocytes in vertebrates (Fig. 12). In summary, I propose that genes in groups A, B, and C likely correspond to the molecular core of the ancestral glial cell in vertebrates. Following the divergence between cyclostomes and gnathostomes, the expression patterns of these genes changed between species, likely due to a cell divergence event in gnathostomes. Genes in group C either lost or reduced their expression in gnathostome astrocytes, resulting in only genes from Groups A and B maintaining their expression pattern in these cells. However, OPCs/oligodendrocytes seemed to preserve the ancestral expression pattern for these genes (group C), akin to lamprey astrocytes. Furthermore, some of the genes in group D are not present in the lamprey genome, so they are unique to gnathostomes (e.g. *Olig1*, *Olig2*, *Mag*), and their expression appears be specific to OPCs/oligodendrocytes, suggesting that their emergence may have been pivotal in oligodendrocyte specification. Notably, only a copy of the gene *Plp1/Gpm6b* is present in the lamprey genome. But this gene underwent a duplication event in gnathostomes. In these animals, *Plp1* is expressed almost exclusively by OPCs/oligodendrocytes (group D), while *Gpm6b* is expressed both glial cells (group B), with higher expression in astrocytes (Fig. 11). Additionally, these cells seem to have reduced expression of genes from group A but retained the expression of genes from groups B and C. Notably, for this model, I did not consider genes whose expression is conserved only among gnathostomes because, with the current species sampling, it is difficult to determine whether these expression patterns were acquired by gnathostome glia or lost in cyclostomes.

This model does not differentiate between OPCs and mature oligodendrocytes, although conserved glia gene expression patterns seem primarily associated with OPCs. This might suggest that during maturation, oligodendrocytes lose the common glial gene expression pattern and switch to more specific genes. However, since this pertains to a developmental process, I considered OPCs as a cell state of oligodendrocytes for this model. Moreover, the difference in gene expression patterns between OPCs and oligodendrocytes likely explains the correlation pattern observed in Figure 9.

Due to the limited number of glial cells recovered from gnathostome species, the proposed hypothesis is confined to a pseudo-qualitative analysis of gene expression levels. To more comprehensively understand the changes in gene expression dynamics that occurred during the evolution of vertebrate glia, additional data, including spatially resolved transcriptomic data, must be acquired.

3.4 Cellular and molecular evolution of the vertebrate telencephalon

In the following section, I focus on findings pertinent to the evolution of the telencephalon, the most anterior part of the brain. This region is associated with higher cognition, as well as various behavioral and sensory adaptations. Furthermore, the telencephalon, as a structured tissue, represents a significant innovation of vertebrates[1]. Our initial goal was to understand its cellular origins by comparing cell and gene expression composition in the telencephali of lampreys and mice. Subsequently, I aimed to explore how the telencephalon evolved within vertebrates. This examination was limited to the cellular evolution of inhibitory and excitatory neurons within the two primary divisions of the telencephalon: the sub-pallium or ventral telencephalon, and the pallium or dorsal telencephalon. A further description of the cell diversity in other regions of the lamprey brain and its implications for vertebrate evolution can be found in Lamanna, Hervas-Sotomayor et al.[5]

3.4.1 Origin and trajectory of inhibitory (GABAergic) neurons in the vertebrate sub-pallium

We identified inhibitory neurons in the lamprey that display a transcriptomic signal characteristic of the medial and lateral ganglionic eminences (MGE and LGE). These structures, located in the sub-pallium, serve as embryonic proliferation zones and are sources of GABAergic neurons. These neurons later migrate to the pallium and the olfactory bulb but also populate adult structures of the sub-pallium[3] (Fig. 3). Prior analyses have hinted at the existence of MGE and LGE in the lamprey brain[8,33,40] . However, no study has yet performed a systematic transcriptomic assessment at the cellular level; therefore, the degree to which these structures are conserved across vertebrates has remained elusive. We focused on studying MGE and LGE-derived cells in adult brains, a subject not comprehensively researched before. Nevertheless, because the MGE and LGE are developmental structures, I begin by providing evidence of their presence in lamprey embryos.

3.4.1.1 MGE and LGE transcriptomic signal in lamprey embryos

The lamprey embryonic brain data showed a cell diversity similar to that of the adult and larval atlases. However, since the embryonic data originated from whole embryo heads, a significant number of cells from non-nervous tissues were recovered, posing challenges for systematic analysis. Consequently, these data were primarily used to confirm the existence of the LGE and MGE zones, with further findings derived from the adult and larval atlases. By analyzing data from two embryonic stages (Tahara stages 26 and 30, Fig. 13; Suppl. Fig. 2), I identified distinct cell clusters expressing genes associated with the MGE and LGE. Notably, transcription factors such as *Meis2*, *Lhx6/8*, and *Nkx2-1/4a* (Fig. 13b, g-h; Suppl. Fig. 2b,f,g) serve as general markers of these structures. Their expression, along with that of effector genes, confirms the presence of these proliferative zones in the lamprey embryonic brain.

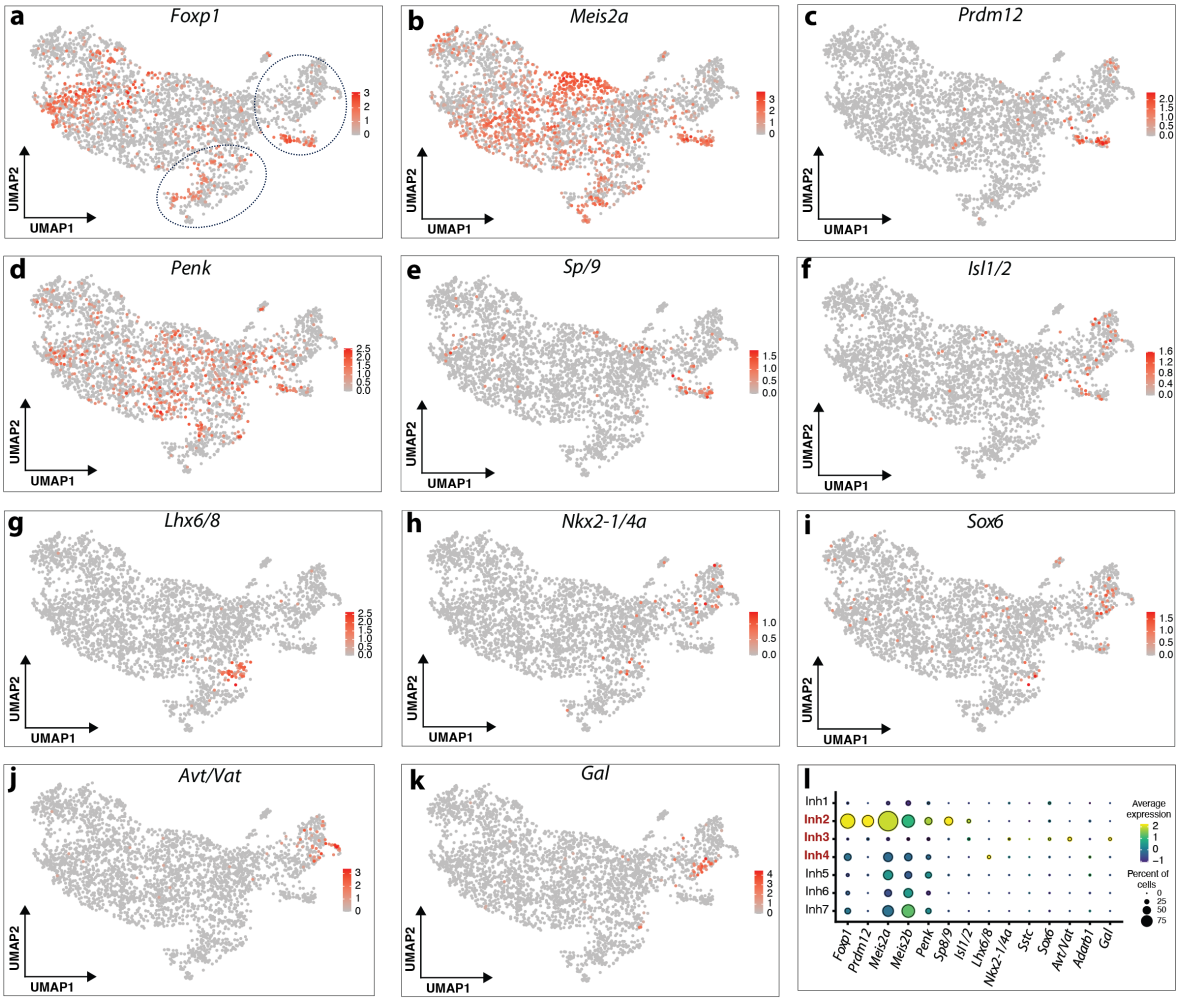


Figure 13. Gene expression in developmental GABAergic cells. a-k) Feature plots showing the expression of genes associated with the medial ganglionic eminence (MGE) and lateral ganglionic eminence (LGE). The dotted circles in (a) indicate the approximate location of these cells based on the concerted expression of all genes. l) Dot plot depicting that three out of seven inhibitory neuron types express the genes presented in a-k. The data correspond to embryos at Tahara stage 30.

Overall, both developmental stages show similar gene expression patterns, but I observed slight differences between them. For example, at the earlier stage (Tahara 26, Suppl. Fig. 2j), cells expressing LGE and MGE specific genes appear to form a single cluster. By the Tahara 30 stage, these cells differentiate into at least three distinct clusters (Inh2-4, Fig. 13l). These observed differences align with developmental expectations: the Tahara 26 stage has been associated with the putative MGE and LGE specification and cell proliferation[33], whereas Tahara 30 corresponds to the hatching stage, by which time cell differentiation and migration are likely underway. However, because of the low count of GABAergic neurons, particularly at stage 26, some of the observed differences between developmental stages may be due to technical limitations.

3.4.1.2 MGE- and LGE-derived cells in the lamprey adult telencephalon

As previously mentioned, we focused further analyses on the adult and larval data to trace the fate of MGE- and LGE-derived cells and to evaluate the conservation of these derivatives across vertebrates. The following results pertain to the adult atlas, though similar findings were described in the larvaes.

By conducting recursive clustering, we identified seven cell types with distinct transcriptomic signatures in the adult telencephalon: two derived the LGE and five derived the MGE (Fig. 14a-b). These cell types displayed clear clustering differences and distinct gene expression patterns, validated spatially by *in situ* sequencing (Fig. 14c-g). In the subpallium, we identified LGE-derived cells in the striatum, while MGE-derived cells were located in two subpallial areas: the subpallial amygdala (SPA) and the pallidum, which is situated caudally to the SPA (Fig. 14e-g). Previously, anatomical labeling recognized the regions now identified as the striatum and SPA as the medial preoptic area (MPO) and striatum, respectively. However, our findings suggest a different transcriptomic identity for these cells. Therefore, we propose that the nomenclature of these structures should be re-evaluated in light of the transcriptomic patterning.

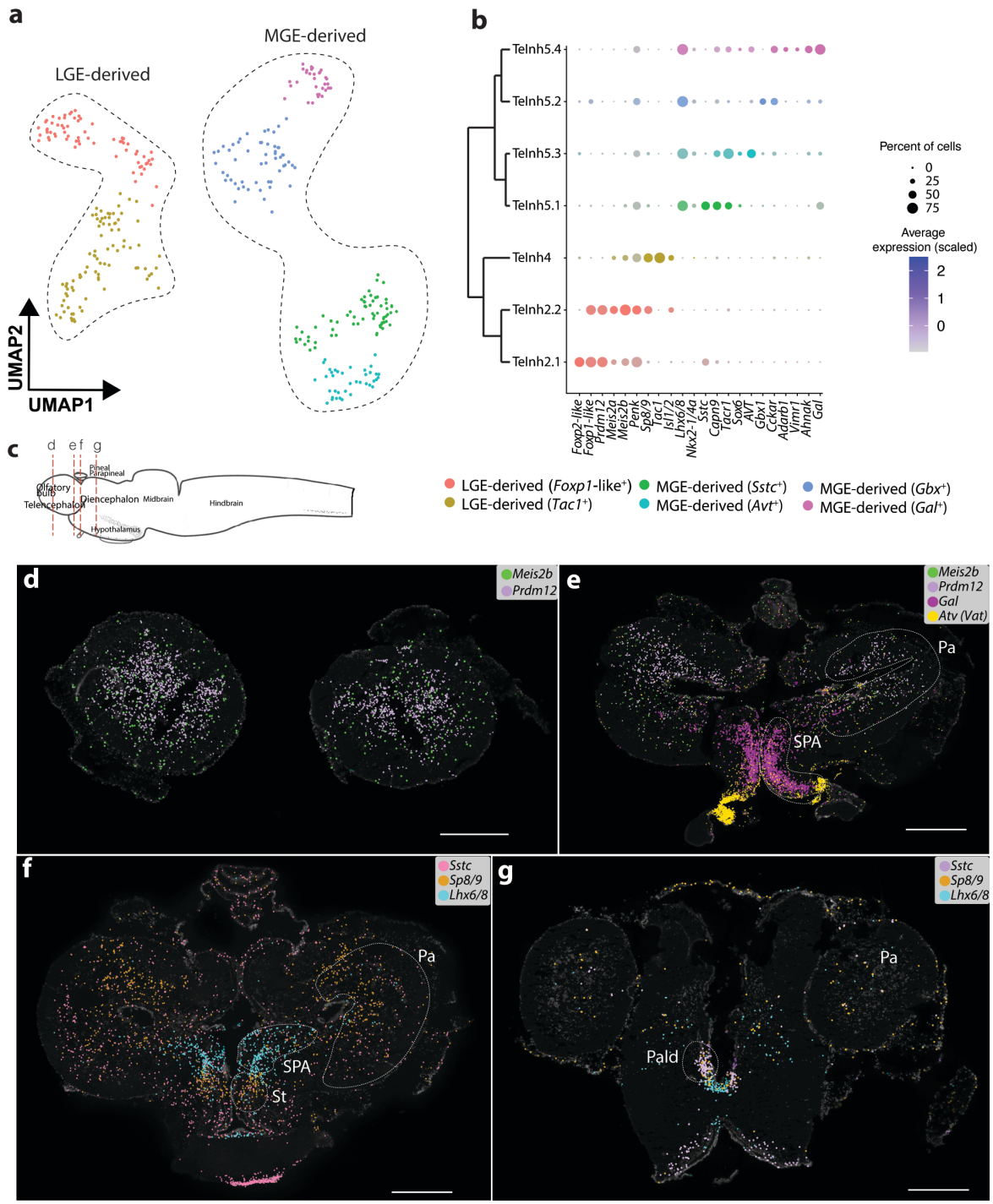


Figure 14. Diversity of GABAergic neurons in the adult lamprey telencephalon. a) UMAP showing LGE and MGE-derived GABAergic neurons. b) Dendrogram depicting the relationships between the clusters identified in (a), along with the expression of selected marker genes. c) Schematic showing the plane of sections in (d-g). d-g) ISS maps of the selected marker genes from (b), across coronal sections through the olfactory bulbs (d), anterior (e, f), and posterior (g) telencephalon. Abbreviations: Pal, pallium; Pald, pallidum; SPA, subpallial amygdala; St, striatum. Scale bars represent 500 µm. *Figures (a-b) were adapted from Lamanna, Hervas-Sotomayor et al*[5]*, license under Creative Commons CC BY.*

Consistent with findings in other vertebrates[48,92,93], MGE-derived cells were also found in the pallium (Fig. 14f, 15), reflecting the presence of GABAergic interneurons (*Sstc*+ cells), while LGE derivatives appeared in the olfactory bulb (Fig. 14d). Intriguingly, I further identified some of these LGE derived cells in the pallium as well (Fig. 14e-f, 15, a pattern that has only been identified before in birds[94–96], though some evidence has also been reported for turtles[97]. However, in other vertebrates studied so far it seems that all pallial GABAergic interneurons are typically MGE-derived[48,93,98].

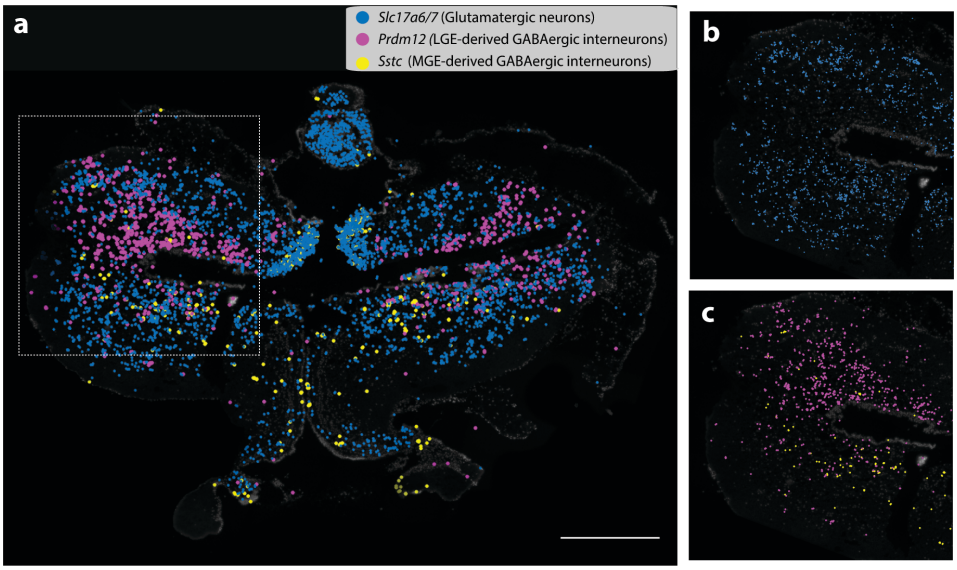


Figure 15. MGE and LGE-derived GABAergic interneurons in the lamprey pallium. a) ISS map across a coronal section of the telencephalon. The white square highlights the area enlarged in b and c. *Slc17a6/7* marks glutamatergic neurons (b). *Prdm12* and *Sstc* mark LGE and MGE-derived interneurons, respectively (c). Refer to Fig. 14c for the section plane. Scale bar represents 500 µm.

3.4.1.3 Conservation of MGE and LGE migration program

The MGE and LGE cells identified in the embryonic stages express *Dlx* genes, transcription factors associated with the tangential migration of MGE- and LGE-derived cells in other vertebrates[48,52,93]. This expression is consistent across various life stages in the lamprey (Fig. 16), suggesting a conserved developmental gene program associated with the migration of these cells that is shared across the vertebrate lineages, despite differences in the locations of LGE-derived interneurons as mentioned above. Altogether, these findings indicate that the proliferative zones of the MGE and LGE, along with their genomic specifications, are conserved across vertebrates and likely existed in their common ancestors.

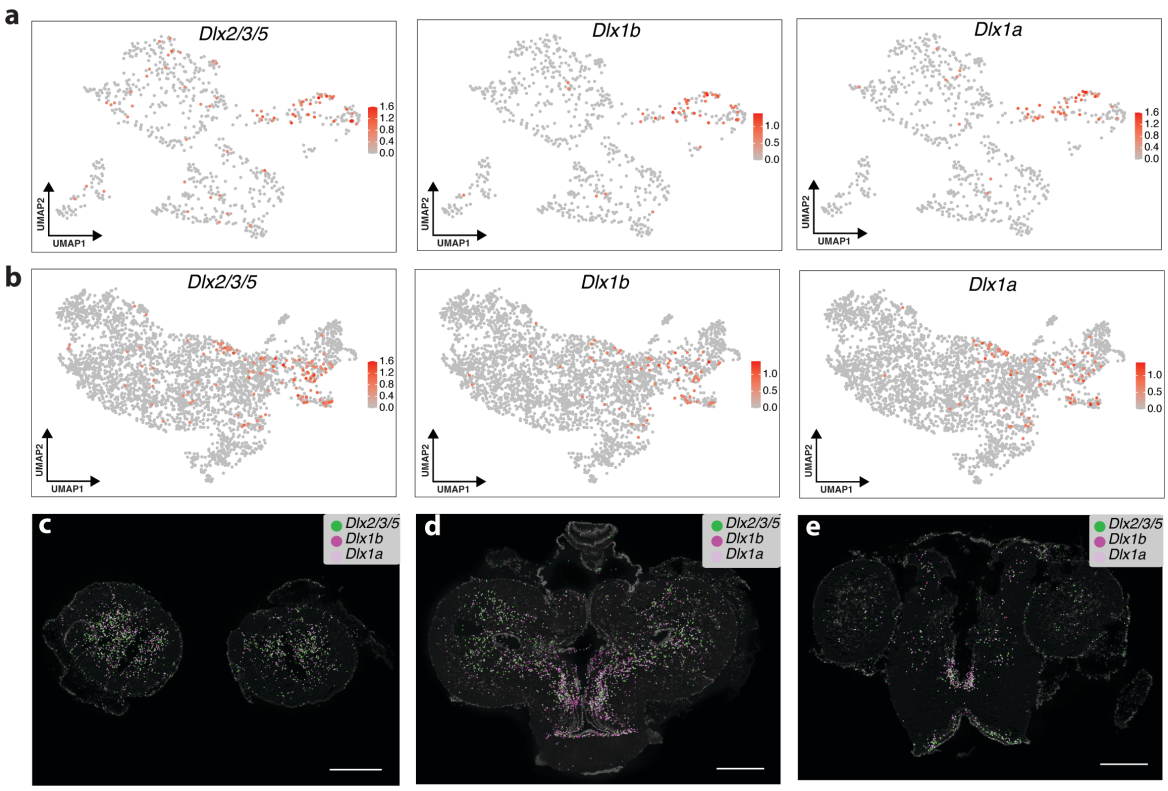


Figure 16. Evidence of tangential migration of GABAergic interneurons in the lamprey telencephalon. a-b) Feature plots illustrating the expression of genes associated with tangential migration of GABAergic interneurons in embryos at Tahara stages 26 (a) and 30 (b). c-e) ISS maps of the genes from a-b, displayed across coronal sections through the adult olfactory bulbs (c), anterior telencephalon (d), and posterior telencephalon (e). The planes of these sections correspond to those in Fig. 14c. Scale bars represent 500 µm

3.4.2 Evolution of excitatory (glutamatergic) neurons in the vertebrate pallium

3.4.2.1 Diversity and organization of the lamprey pallial neurons

We identified eight distinct excitatory cell types in the lamprey pallium (Fig. 17a-b). These cell types exhibited striking similarities to the excitatory cells found in the mouse telencephalon. The identified cell types were categorized into four main regions: the dorsomedial telencephalic nucleus (DMTN, TeExc1), the anterior pre-thalamic eminence (PthE, TeExc4, TeExc2), the pallial extended amygdala (PEA, TeExc3), and the rest of the pallial excitatory neurons (TeExc2 and TeExc5–8) (Fig. 17a-e). Interestingly, because of their localization, this result confirmed the presence of PThE neurons in a region of the lamprey brain traditionally identified as medial pallium[56], agreeing with previous molecular reports[41,67]. DMTN cells, on the other hand, displayed a transcriptomic signature akin to that of the mitral cells or tufted

cells of the olfactory bulb, positioned at the interface between the pallium and the olfactory bulb (Fig.17c-h).

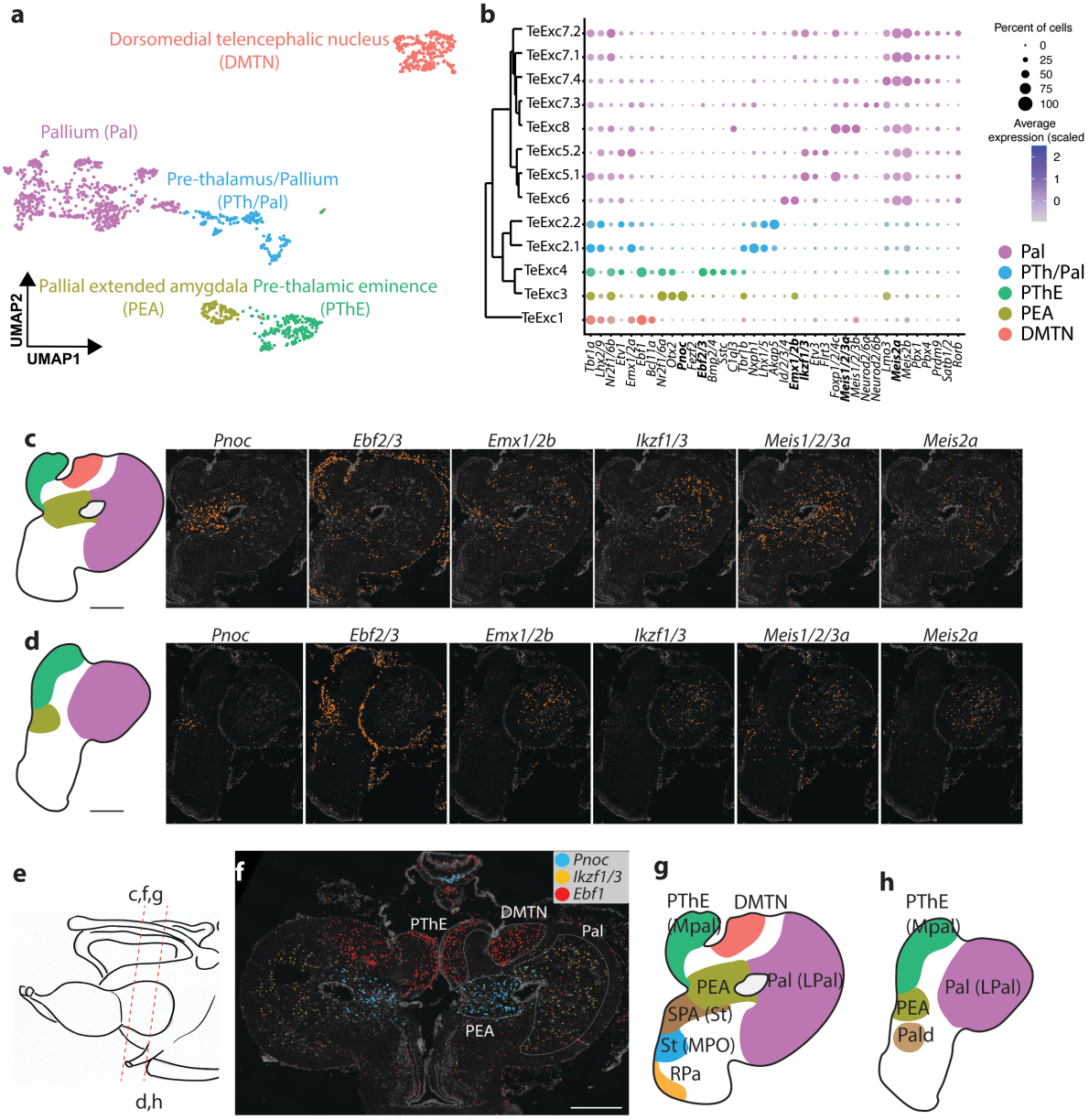


Figure 17. Diversity of excitatory neurons in the lamprey adult telencephalon. a) UMAP of excitatory neurons in the telencephalon, delineating various regions. b) Dendrogram illustrating the relationships between the clusters shown in (a) and the expression of selected marker genes. c-d) Spatial localization of the clusters from (a) and (b), along with the ISS maps of selected marker genes (emphasized in bold within b) in the caudal (c) and rostral (d) coronal sections of the telencephalon. e) A schematic of the sections for c, d, f, g,h. f) The spatial expression (ISS) of selected genes marking regions delineated in (a). g-h) Interpretation of lamprey telencephalic anatomical regions based on this work's findings; the names in parentheses indicate the original designation of the region before this study. Abbreviations: MP: medial pallium, DMTM, dorsomedial telencephalic nucleus; LPal: lateral pallium, PTh, pre-thalamus; Pal, pallium; PEA, pallial extended amygdala, Rpa: rostral

paraventricular area, SPA: subpallial amygdala, St: striatum, Pald: pallidum. Scale bars represent 500 µm. *Adapted from Lamanna, Hervas-Sotomayor et al, license under Creative Commons CC BY* [5]

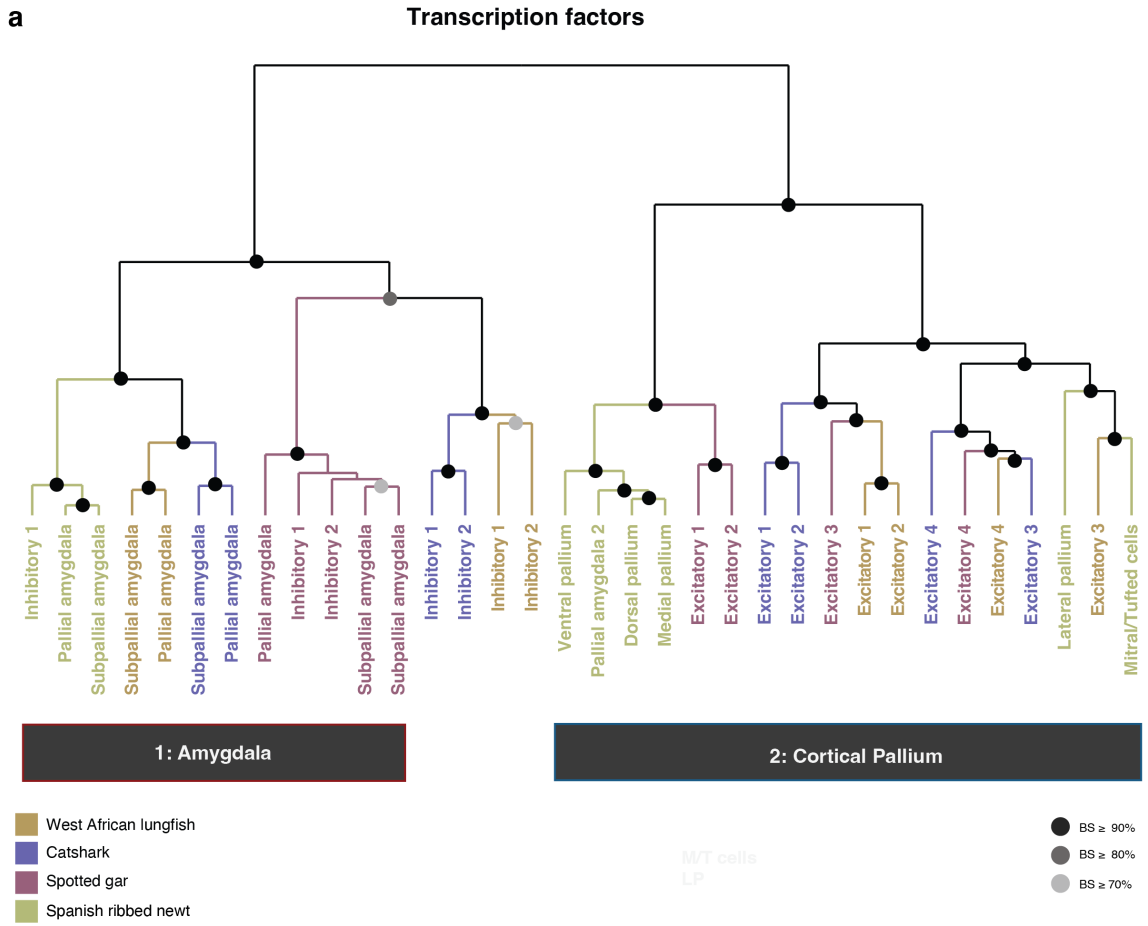
Despite displaying a specific gene expression profile that distinguishes them based on the clustering analyses of the single cell transcriptomic datasets, the pallial neurons in lampreys do not seem to be organized into spatially distinct regions (Fig. 17c-h). In contrast to tetrapods, the lamprey pallium appears to consist of a homogeneous domain that comprises at least four transcriptomically distinct cell types ((TeExc2 and TeExc5–8, Fig. 17b). This indicates that differentiation into spatially distinct pallial regions (such as lateral, ventral, dorsal, and medial) probably evolved after the cyclostome-gnathostome split. This finding contrasts with previous assumptions that the regionalization based on connectivity data would reflect directly in the transcriptomic layer. Additionally, it challenges the homologization of some of these regions with particular areas of the mammalian pallium[55–58].

3.4.2.2 Excitatory neuron clades across gnathostome species

To further investigate the evolution of the pallium in vertebrates, I examined the correlation of pallial cells among gnathostomes. I performed a comparative analysis across various species, including the catshark, spotted gar, lungfish, and a salamander species with well-characterized pallial regions at the cellular level[48]. My approach used "pseudobulk" analysis to infer average gene expression for each defined cell group, correlating cells across species. Due to the limited number of one-to-one orthologues between these phylogenetically distant species, I estimated average expression for orthogroups—sets of genes with high sequence similarity across multiple species, likely descending from a single gene in the last common ancestor (details in the methods section). I also included a set of sub-pallial inhibitory neurons from each species to serve as an "outgroup," helping to infer correlations among excitatory neurons. These analyses correspond to the initial exploration conducted on these data. Note that the excitatory cells from catshark, gar, and lungfish have not yet been functionally annotated, primarily because the expression of canonical excitatory neuron markers does not follow the same patterns as in amniotes, as discussed below. Consequently, the use of these genes does not yield significant information for a detailed annotation of these cell groups and cell types. These correlations were intended to assist in the annotation process by considering shared expression patterns across a large set of genes, while simultaneously shedding light on the diversification of these cells.

In the subsequent sections, I make reference to the nomenclature from the tetrapartite model (medial, dorsal, ventral and lateral pallium)[45]. The use of this terminology serves practical purposes; although the model seeks to elucidate the evolutionary relationships of pallial regions in amniotes, its application to anamniote species is not straightforward. Additionally, this analysis relied on the brain cell annotation in salamanders by Woych et al[48], performed based on comprehensive spatial and quantitative transcriptomics. As a result, it contains the following deviations from the existing tretrapartite model: 1) the area identified as the lateral pallium corresponds to a zone of olfactory-recipient cells, unlike the defined region in mammals, which includes the claustrum and insular cortex; 2) due to the cell-focused nature of this analysis, a specific cell terminology will be used. Neurons previously classified under the pallial amygdala will be considered independently. In contrast, neurons from other pallial divisions (medial, dorsal, ventral, and lateral) will be collectively termed 'cortical'. This cellular categorization aligns with the terminology used by Woych et al.[48]

I correlated cell clusters using the whole transcriptome and only transcription factors, as the latter are generally more conserved in expression and considered to be the defining features of cell types[71]. Using both gene sets, I consistently identified at least two main excitatory neuron clades, named after the salamander cells that they include: the amygdala and cortical pallium (Fig. 18, Suppl. Fig.3-4, see paragraph above for the cell nomenclature description). While there are differences concerning the internal nodes within the cortical pallium clade, the dendrograms demonstrate consistency in the internal relationships within the amygdala. Across all species examined, excitatory cells from the amygdala were uniformly more closely related to inhibitory cells within the amygdala relative to other excitatory neurons, forming a distinct clade in the dendrogram. The amygdala of the gar showed a closer correlation with other gar inhibitory neurons rather than with the amygdalar cells of other species. This pattern could result from a stronger phylogenetic signal in gar cells or pronounced differences in the gar's amygdala. Notably, one cell type from the salamander amygdala did not cluster with this clade but rather with the cortical pallium. Altogether, these findings suggest a shared origin of the amygdalar cells for these species with subsequent cell diversification that may be unique to certain lineages. Additionally, due to the lack of a direct correlation between numerous amygdala cells and the ventral pallium, it is conceivable that, at least in salamanders, these cells do not share an association or a common origin, which contrasts with the traditional consideration of the amygdala as part of the ventral pallium, under the tetrapartite model[99].


a
Transcription factors
Inhibitory 1
Pallial amygdala
Subpallial amygdala
Subpallial amygdala
Pallial amygdala
Subpallial amygdala
Pallial amygdala
Pallial amygdala
Inhibitory 1
Inhibitory 2
Subpallial amygdala
Subpallial amygdala
Inhibitory 1
Inhibitory 2
Inhibitory 1
Inhibitory 2
Ventral pallium
Pallial amygdala 2
Dorsal pallium
Medial pallium
Excitatory 1
Excitatory 2
Excitatory 1
Excitatory 2
Excitatory 3
Excitatory 1
Excitatory 2
Excitatory 4
Excitatory 4
Excitatory 4
Excitatory 3
Lateral pallium
Excitatory 3
Mitral/Tufted cells
1: Amygdala
2: Cortical Pallium
West African lungfish
Catshark
Spotted gar
Spanish ribbed newt
BS ≥ 90%
BS ≥ 80%
BS ≥ 70%

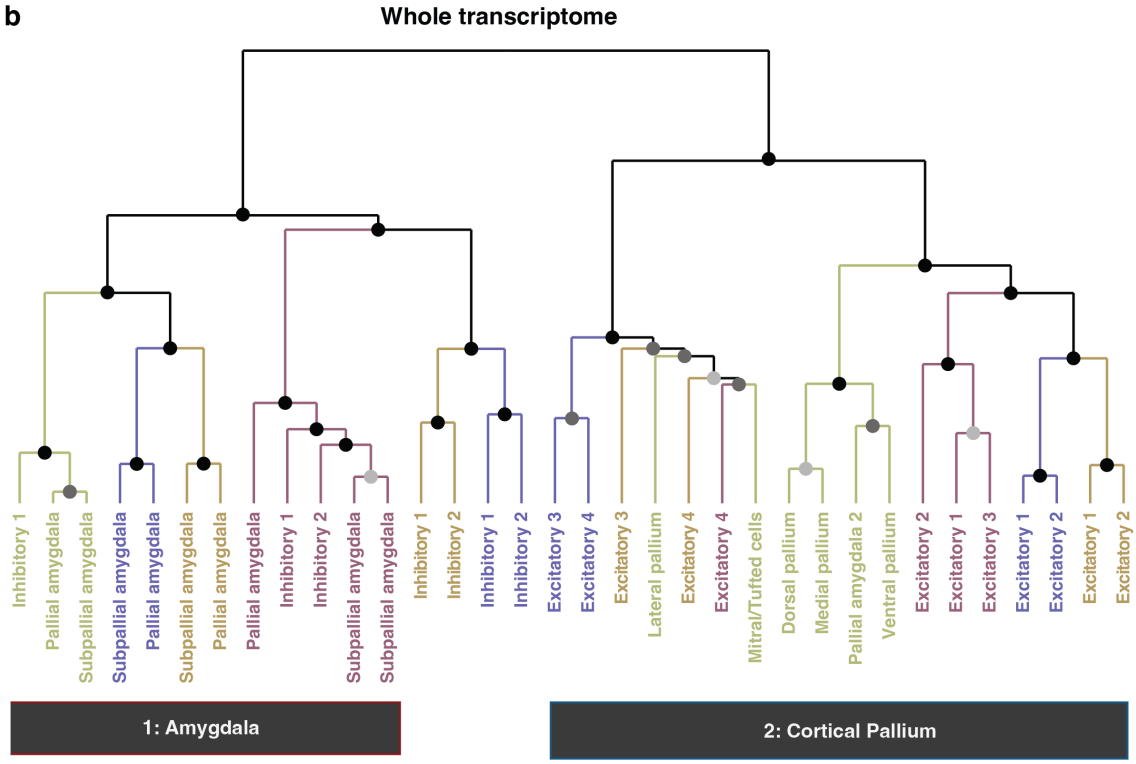

b
Whole transcriptome
Inhibitory 1
Pallial amygdala
Subpallial amygdala
Subpallial amygdala
Pallial amygdala
Subpallial amygdala
Pallial amygdala
Pallial amygdala
Inhibitory 1
Inhibitory 2
Subpallial amygdala
Subpallial amygdala
Inhibitory 1
Inhibitory 2
Inhibitory 1
Inhibitory 2
Excitatory 3
Excitatory 4
Excitatory 3
Lateral pallium
Excitatory 4
Excitatory 4
Mitral/Tufted cells
Dorsal pallium
Medial pallium
Pallial amygdala 2
Ventral pallium
Excitatory 2
Excitatory 1
Excitatory 3
Excitatory 1
Excitatory 2
Excitatory 1
Excitatory 2
1: Amygdala
2: Cortical Pallium

Figure 18. Comparisons of telencephalic excitatory neurons in gnathostomes. Dendrograms display the gene expression distance (Spearman's ρ) for transcription factors (136 orthogroups) and b) whole transcriptome (4,546 orthogroups). The analysis clustered excitatory cells into two main clades, labeled according to the inclusion of a corresponding salamander cell class. Supp. Fig. 4 presents the same analysis based on downsampled data (500 cells per cluster). Dendrograms' overall topology—based on the cells clustered within these two clades—is consistent across all analyses. Internal nodes are different between dendrograms. Different colors denote distinct species. Approximately unbiased (au) bootstrap support (BS) is provided (n = 1,000).

In the other cell clade, it is notable that the salamander lateral pallium does not correlate with the rest of the salamander's cortical neurons but rather with the tufted and mitral cells derived from the olfactory bulbs. This correlation suggests a transcriptomic association, likely due to their role in olfaction perception, as mitral and tufted cells are projection neurons that extend into the olfactory cortex of the pallium. Alternatively, this correlation could result from an statistical artifact (akin to long branch attraction[100]), meaning that because these groups of cells are transcriptomically distinct from the rest, they correlate. Further inspection of these cells will elucidate the true scenario. However, the primary finding is that these cells do not seem closely associated with other cortical cells from the salamander brain; in fact, they don't even cluster in the same cell type subclade. This observation suggests that the division of cortical excitatory cells into at least two subclades (medial, dorsal, ventral, vs. lateral) may be ancestral to gnathostomes. Additionally, this finding indicates that a region of olfactory recipient cells (here as lateral pallium) may have been a substructure already present in the last common ancestor of gnathostomes, as proposed by models of pallial evolution based on neuron projection data[27].

The remaining cortical excitatory neurons of the salamander form a distinct clade that is species-specific in both the transcription factor and whole transcriptome dendrograms, indicating that this diversification is unique for salamanders. However, this particular cell clade collectively has a close transcription factor correlation with a cell clade from the gar (Fig. 18a, Suppl. Fig.3-4). Correlations based on the whole transcriptome dendrogram, however, suggest the existence of an interspecific cell clade for cortical neurons across gnathostomes (Fig. 18b, Suppl. Fig.3-4). Nonetheless, both dendrograms seem to indicate the absence of one-to-one relationships between cortical excitatory neurons across species. However, comparing the data at various levels of clustering resolution might uncover such relationships. Thus, further detailed characterization of non-tetrapod pallia will help in clarifying these patterns of correlation.

To further understand the evolution of pallial cells in these species, I explored the expression of canonical transcription factors associated with specific pallial regions in tetrapods[47,48,69] (Fig. 19a). While transcriptomic signatures are recognizable across species, the distinctive expression profiles per cell class are not as pronounced as those observed in salamander cells, where a clear, concerted expression pattern is defined.

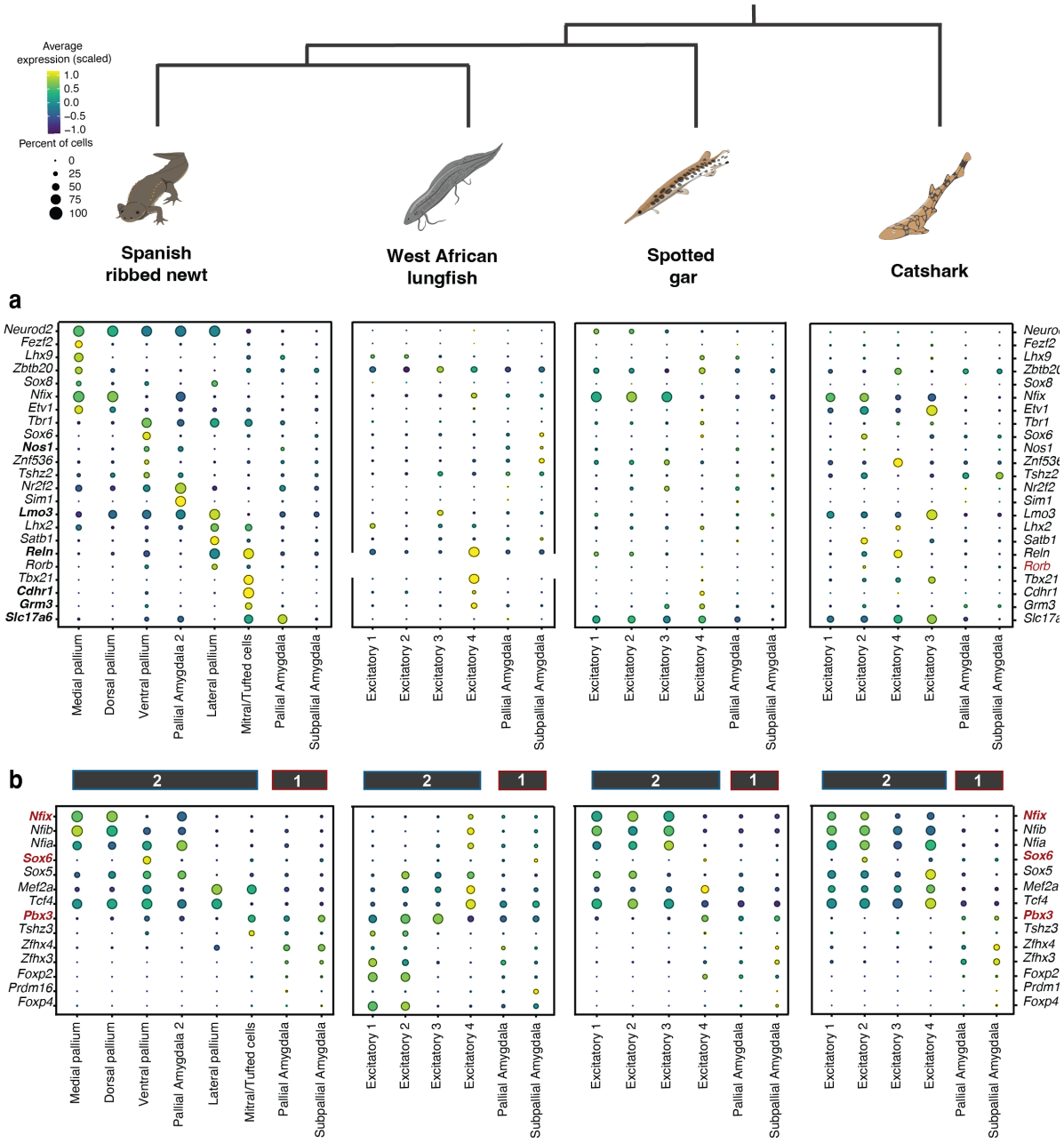


Figure 19. Expression profiles of pallial-associated genes in gnathostomes. a) Canonical genes traditionally used to identify pallial regions in tetrapods, as referenced by Woych et al. Effector genes are in bold. *Rorb* is not expressed in lungfish cells. b) Differentially expressed transcription factors between cell classes that are common for all species. The genes marked in red are transcription factors

associated with ventral (and lateral) or dorsal (and medial) pallial specification. The numbered boxes correspond to the cell clades identified in Fig. 18.

I further assessed the expression conservation of common transcription factors among the four species (Fig. 19b). Specifically, I estimated the top 200 differentially expressed genes between cells within each species (Suppl. Fig.5) and from those genes, I selected the transcription factors common for all species. This analysis uncovered a lightly more defined transcriptomic signal consistent with the previous results, depicting higher conservation between catshark, spotted gar, and salamander. Notably, in these three species there is a similar expression pattern of *Nfix*, *Pbx3* and *Sox6,* genes associated with dorsal and ventral fates in the salamander embryonic pallium, respectively[48]. In contrast, lungfish displayed an inverse expression pattern for these genes. Still, the lungfish presents a more distinct transcriptomic signature between cell classes than the other two species (Suppl. Fig.5). Nevertheless, despite some level of homogeneity in catshark and spotted gar, excitatory neurons across all three species generally exhibit a cell class-specific transcription factor expression signature. These expression patterns together with the correlations suggest that broad cell clades (amygdala and cortical cells) may share a common origin in the ancestor of these species. Additionally, a distinct region of olfactory-receptor cells may have been present in the gnathostome ancestor. Cells examined at a higher resolution, however, appear to be lineage-specific, a result of cell diversification events that possibly occurred after the lineages diverged.

4 Discussion

4.1 Comparative analysis of cell type evolution in the vertebrate brain

Ongoing efforts to understand cell evolution and diversification have intensified with the advent of single-cell RNA sequencing technologies. As multiple cell atlases become available, so do attempts to compare them between species and gain insights into cellular evolution. However, in vertebrates, the focus is predominantly on amniotes, especially mammals[14,23,45,99], leading to a sampling bias in evolutionary studies of tissues like the brain, which originated before mammals emerged. The cell diversity, structural complexity, and adaptive significance of the vertebrate brain have long fascinated scientists, but these very features have also complicated its study. An integrative approach, extending beyond mammalian species, is required to address its evolution. This study is a pioneer in comparative cell evolution analysis among vertebrates, focusing on lineages predating amniotes and substantially contributing to bridging these aforementioned research gaps.

Overall, the results obtained from lamprey-mouse comparisons increase our understanding of the molecular and cellular origins of the vertebrate brain. Insights from the exploration of the other vertebrate species further contribute to our comprehension of cell diversification throughout ~600 million years of vertebrate evolution. Additionally, this provides light on how homologous cell type families have diversified after the divergence of vertebrate lineages, and offers clues on the possible interplay of cell and tissue evolution.

Altogether, these analyses have identified cell types, tissues, and molecular programs conserved across vertebrates, which likely correspond to the ancestral core of the vertebrate brain (Fig. 20). For example, the correspondence of broad cell type classes and families among vertebrate species, along with the conserved, concerted expression profiles of transcription factors, suggests that deep homologies exist within these hierarchical ranks. However, establishing correspondences at finer resolutions (e.g. cell types) is more challenging because one-to-one cell type homologies are unexpected between phylogenetically distant species—a result reported by previous studies comparing cortical neurons among amniotes[52,70] . This is a consequence of the lineage and cell diversification process itself; while homologies between cell type families or classes can be identified across distant lineages, divergence leads to the accumulation of differences. This phenomenon mirrors the evolutionary progression of

species, where the demarcation between closely related species tends to be blurry[72]. However, as evolutionary units diverge over time, their distinctions tend to become more marked. As a result, differentiating between animal families or orders is more straightforward than between closely related species.

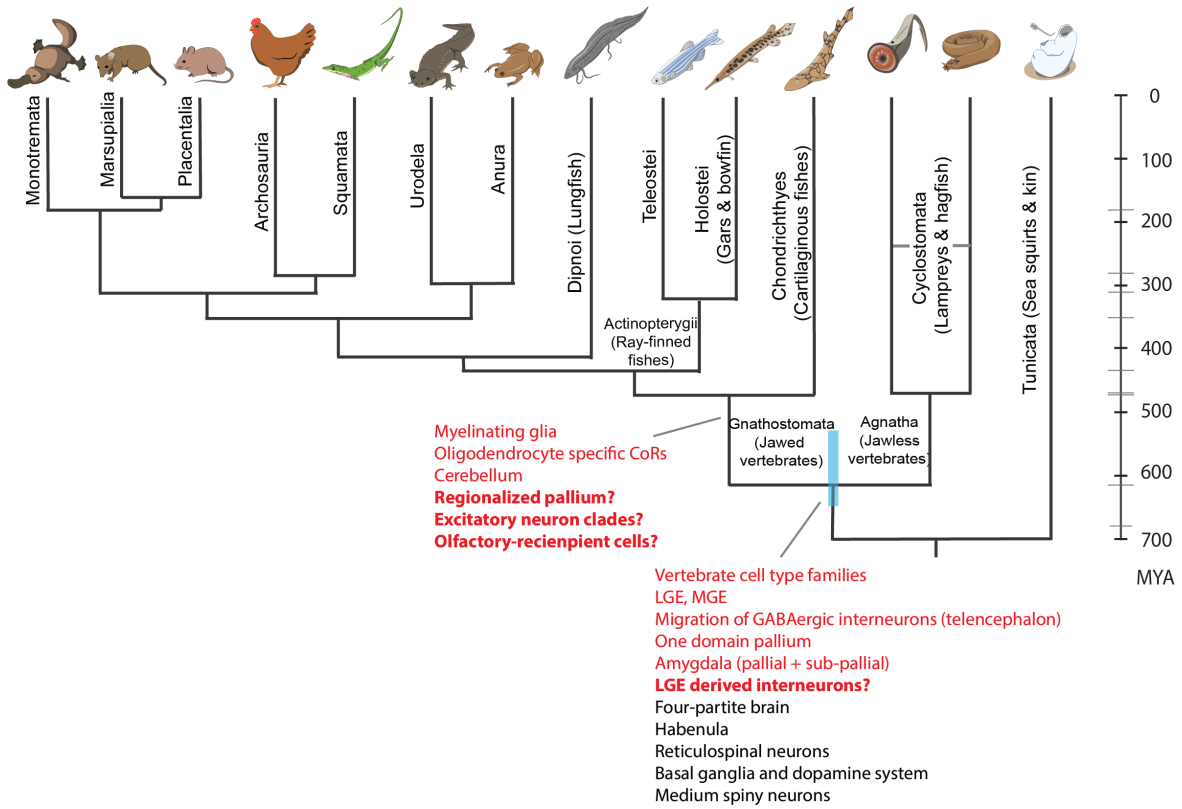


Figure 20. Brain innovations in the vertebrate and gnathostome ancestors. Vertebrate phylogenetic tree as in Fig. 1a, illustrating significant brain innovations (highlighted in red) as indicated, shown or confirmed by this thesis. Resulting open questions from this work are emphasized in bold. Findings not mentioned here but discussed in Lamanna, Hervas-Sotomayor et al[5]. are depicted in black.

4.2 Diversification of glial cells and the origins of myelination

My observation of conserved expression patterns of transcription factors aligns with the cell genetic individualization process, where effector genes become differentiated but transcription factors remain conserved[71,76]. This phenomenon is exemplified by the results from glial and pallial cells, which are broad classes homologous across vertebrates but contain cell types with unique gene expression programs in each lineage. In this context, glial cells appear to have undergone a dynamic evolutionary history characterized by cell divergence, associated with gene duplication (Fig. 12). For instance, the identification of astrocytes in lampreys that express

several oligodendrocyte-specific genes, including transcription factors, indicates an evolutionary connection between these cell types. These insights complement previous studies, which indicated that lamprey axons seem to be physically associated with astrocytes[91] and that some elements of the regulatory program essential for oligodendrocyte differentiation are expressed during lamprey gliogenesis[42]. Moreover, my comparative analysis across vertebrates suggested that oligodendrocytes likely evolved from an ancestral precursor cell that underwent duplication and subsequent divergence in the gnathostome lineage (Fig. 12). The shared expression program between lamprey astrocytes and gnathostome glia (astrocytes and oligodendrocytes) implies that the ancestral cell was already equipped with part of the molecular toolkit necessary for oligodendrocyte development. However, it is significant that other key components of this molecular toolkit are genes unique to gnathostomes, suggesting that the emergence of oligodendrocytes may be linked to the genome duplication event that occurred within this lineage[101].

4.3 Role of gene duplication and paralogue switching in vertebrate cell diversification

My results provide insights into the connection between gene duplication and neofunctionalization, leading to the differential expression of paralogues and the potential for genetic individualization in cells. For instance, the analyses indicate that the expression of the paralogues *Slc17a7* (*Vglut1*) and *Slc17a6* (*Vglut2*) has diverged during gnathostome evolution. The broad expression of the single gene copy across all lamprey excitatory neurons suggests that *Slc17a6/7* has an ancestral role in vertebrate glutamatergic cells. However, in gnathostomes, this gene underwent a duplication event whereafter each paralogue has adopted distinct expression patterns in different lineages. For instance, in mammals, their expression seems to be complementary with minimal overlapping, being *Slc17a7* the canonical telencephalic marker [84,86]. In the shark's telencephalon, both genes mark excitatory neurons, similarly to the co-expression of these genes in turtles[52]. The gar telencephalic excitatory neurons seem to express uniquely *Slc17a6*. Conversely, in the lungfish, *Slc17a7* is the general glutamatergic marker in the telencephalon. This pattern is similar in the salamander telencephalon, where *Slc17a6* is exclusively expressed in the glutamatergic cells of the amygdala[48,82]. These observations indicate that regulation of these genes has undergone multiple changes throughout vertebrate evolution. Given the complementary expression of these genes in the mammalian brain is proposed to be linked to functional differences in neuronal synapses[86], future comparative functional studies across vertebrates could elucidate

whether these changes correlate with cell phenotypes. Moreover, variations in the expression levels of these genes have been observed in human patients with brain-associated diseases such as schizophrenia and Parkinson's disease[86,102], so analyzing the regulatory changes during evolution could provide valuable information for understanding the genetic basis of these disorders.

A more pronounced example is given by *Plp1* and *Gpm6b*, paralogues resulting from another gene duplication event in the ancestor of gnathostomes. The expression of the single copy orthologue across lamprey astrocytes suggests that this gene was expressed in this glial class in the shared ancestor of vertebrates. In gnathostomes, however, each paralogue has restricted its expression: *Gpm6b* is expressed by astrocytes and oligodendrocytes (albeit the expression is higher in the former), and *Plp1* by oligodendrocytes. These observations align with other single-cell studies that identified paralogue substitutions among homologous cell types in different animal clades[77,103].

4.4 Evolutionary origins of inhibitory neurons in the vertebrate sub-pallium

The identification of inhibitory neurons, derived from both the lateral and medial ganglionic eminences (LGE and MGE), in the embryonic, larval, and adult telencephalon of lampreys, indicates the presence of two primary GABAergic progenitor zones in the subpallium of the common vertebrate ancestor, as previously suggested[33,104]. However, these findings challenge the traditional views on lamprey neuroanatomy, especially the localization of the main subpallial regions (Fig. 17g-h). The spatial distribution of cells derived from the LGE and MGE suggests that the region traditionally identified as the medial preoptic nucleus (MPO) in lampreys may indeed be homologous to the dorsal striatum of gnathostomes, and should thus be reclassified. Furthermore, the region traditionally associated with the striatum in lampreys exhibits marker expression patterns that align with those found in the subpallial amygdala of gnathostomes. For a detailed description of this reclassification, see the discussion in Lamanna, Hervas-Sotomayor et al.[5].

Furthermore, these results suggest that, as in other vertebrates, neurons derived from the LGE and MGE contribute to the populations of GABAergic interneurons in the olfactory bulbs and pallium, indicating some conserved migratory patterns across vertebrate species. Notably, however, lamprey pallial interneurons arise from both the MGE and LGE (Fig. 15), which

differs from patterns observed in the pallia of mammals[93] and amphibians[48]. Typically, in these species, all pallial GABAergic interneurons originate from the MGE, the CGE, and, in mice, the preoptic area, whereas LGE-derived interneurons are restricted to the olfactory bulbs[48,93,98]. Conversely, numerous studies have pointed to the presence of LGE-derived interneurons within the pallium of birds[94–96], and there is some evidence suggesting their existence in turtles as well[97]. These findings suggest that, while subpallial structures are conserved across vertebrates, the derived interneurons that migrate to the pallium have different origins. It remains to be clarified whether the exclusive migration of LGE derivatives to the olfactory bulbs is an ancestral trait that changed in some lineages such as lampreys and birds. Alternatively, this could be a derived condition in some gnathostomes, while the ancestral pallium contained interneurons from both MGE and LGE origin. Further comparative research on additional non-tetrapod species will be crucial for understanding the evolutionary history of these cells.

4.5 Evolutionary dynamics of excitatory neurons in the vertebrate pallium

The evolution of the pallium remains one of the most debated and intriguing topics in comparative neuroscience. The controversies result from the complex evolutionary dynamics of this brain region, which is evident in its structural variations across vertebrate species[1]. Recent studies have uncovered a significant divergence in cell types within the pallium, showing higher levels of species-specific variations compared to other neuronal populations, such as those found in the sub-pallium[48,52,70,85].

Accordingly, various evolutionary models have been proposed to compare the pallium across vertebrate lineages[47]. Traditional frameworks, such as the tetrapartite model[23,45], originally formulated for amniotes—particularly mammals—encounter considerable challenges when applied to other vertebrates. Recent molecular and cellular evidence from salamanders confirmed the presence of pallial regions in tetrapods, although direct homology with mammalian pallial regions has not been established[48]. However, beyond this study, most research does not take into account the distinctive brain structures found in anamniotes. Still, if the tetrapartite model is applicable to other anamniotes, as several authors have suggested[1,57,66], it would mean that at least some pallial regions predate the rise of tetrapods. Yet, despite the vast diversity of vertebrates being represented by non-tetrapods, their inclusion in studies of brain cellular evolution has been minimal, resulting in a significant phylogenetic

gap in the field. This gap likely contributes to the debate, possibly leading to misinterpretations of features unique to specific evolutionary lineages. My thesis findings narrow this sampling gap by focusing on diverse species from non-tetrapod lineages, enabling a progressive reconstruction of the pallium's evolutionary history. Overall, my findings challenge the concept of direct one-to-one homologies between the four pallial regions of tetrapod and non-tetrapod vertebrates.

4.5.1 The one-domain pallium of the lamprey

The findings in the lamprey dorsal telencephalon have confirmed the existence of mitral-like cells (DMTM), pallial neurons, and pallial amygdala (PEA). Contrary to previous assertions[56], these findings do not corroborate the existence of a medial pallium in lampreys, as the cells in question show a transcriptomic signature indicative of a diencephalic origin (for further discussion, see section 3.4.2.1). Additionally, while the identified pallial neurons in lampreys are transcriptomically distinct, spatial data revealed that they do not localize in distinct regions of the pallium (Fig. 17g-h). These insights suggest that the evaginated region of the lamprey telencephalon, previously identified as lateral or dorsal pallium[55–57,105], should be recognized as a one-domain pallium. This region exhibits cellular expression signatures that are homologous to all pallial subdivisions recognized in tetrapods.

4.5.2 Origin and diversification of pallial divisions in gnathostomes

The previous insights raise new questions regarding the evolutionary emergence of a truly regionalized pallium and its conservation across vertebrates. My comparative analysis between gnathostome species indicates a structured organization of pallial excitatory cells within this lineage (Fig. 18). It seems likely that a division of excitatory neurons in at least two cell clades: pallial amygdala and pallial cortical neurons, was present in the common ancestor of gnathostomes (for a description of the cell nomenclature see section 3.4.2.2). Within the cortical clade, a further ancestral division might have existed, distinguishing cells associated with olfactory perception from other cortical neurons.

The correlation of mitral and tufted cells, which are located outside the pallium in the olfactory bulbs but project into it, with the lateral pallium of salamanders, is consistent with the original anatomical definition of amphibian lateral pallium[106]. This finding further supports the

transcriptomic classification of these cells and their status as homologues of the olfactory-recipient cells of the reptilian lateral cortex[48], and not of the cells identified as lateral pallium in mammals [45,99] (refer to section 3.4.2.2. for further discussion). These results may suggest that the cells associated with the olfactory recipient area of the pallium have an ancient origin, potentially tracing back to the common ancestor of all gnathostomes, or perhaps even all vertebrates, as evidenced by the presence of mitral-like cells in the lamprey pallium. Further spatial data and analyses are needed to confirm the association of these cells with olfactory projections in vertebrates and their relationship to the one-domain pallium of lampreys, which predominantly receives olfactory input[56].

Additionally, my findings provide insights into the origins of the vertebrate amygdala, suggesting that, at least some cells from this structure are ancestral. However, the complexity and regionalization of the tetrapod amygdala, as evidenced by its diversity of cell types and developmental origins[82,107], indicate a dynamic evolutionary history. Additionally, the results from the correlation analyses align with recent findings that propose the amygdala be recognized as a distinct pallial region[47,99,107], contrasting its traditional classification as part of the ventral pallium[45]. Still, certain cells in the salamander amygdala have a transcriptomic profile akin to that of the ventral pallium. These cells were identified as molecularly distinct from other amygdalar cells and initially misclassified as cortical pallium by Woych et al.[48]. However, they were later confirmed to be part of the amygdala based on spatial gene expression patterns. Such cellular diversity supports the proposed notion that the amygdala may be an example of mosaic evolution, where traits develop at different rates due to varied selective pressures[82,108], or that its cell types emerged at different times during vertebrate evolution.

Finally, for the rest of the pallium, the analyses yielded puzzling results. While there is a clear indication of homology among the broader cortical neuron clades across species, the existence of one-to-one homologies remains questionable. This is evidenced by the inconsistent transcription factor profiles across species, despite a collective similarity when comparing broad cell clades between species. It appears that glutamatergic neurons in vertebrates have undergone lineage-specific diversification, leading to distinct transcription factor expression profiles. This observation is consistent with findings in the dorsal pallium of amniotes, where the same transcription factors are collectively expressed across the entire structure but display different cellular combinations in reptiles and mammals[52,70]. It is plausible that the ancestral vertebrate and gnathostome pallium comprised a set of cell classes that diversified along each

lineage, retaining an ancestral regulatory toolkit but manifesting it in varied combinations. Future analyses including a finer cell annotation and spatial data will be crucial for determining whether these distinct cellular trajectories correspond to pallial regions.

Collectively, these findings suggest that at least two major clades of excitatory neurons in the pallium may have emerged before the divergence of gnathostomes. Since similar patterns are observable in lampreys, it's plausible that such neuronal division existed in the last common ancestor of vertebrates. Following the cyclostome-gnathostome split, perhaps another cell clade emerged that separated neurons populating the olfactory recipient area of the pallium. As gnathostomes diverged, lineage-specific cell diversification likely took place within each cell clade, which explains the lack of a direct correlation of cell classes between species, leading to the emergence of many-to-many relationships. Thus, while the pallium overall might be homologous across these species, specific subdivisions within it could be dependent on the lineage. Yet, further cell type characterization within gnathostomes is required to gain a deeper understanding of pallial evolution..

5 Conclusions and Outlook

Altogether, my findings provide valuable insights into the cellular and molecular evolution of the vertebrate brain, including the identification of ancestral structures including cell type families and the developmental sources of GABAergic neurons in the telencephalon. Additionally, these findings shed light on the emergence of gnathostome-specific tissues, cell types and brain substructures, such as myelinating glia, and a regionalized pallium (Fig. 20). Moreover, this research has provided clarity on one of the most debated topics in vertebrate neuroscience—the origins of the cerebellum. Our data suggest that the cerebellum is truly an innovation of gnathostomes. By providing an overview of the ancestral brain at cellular resolution, this investigation creates a foundational framework for further explorations into the cellular diversification and structural evolution across the vertebrate clade.

Extending this project to include gnathostome species significantly contributes to the comprehension of the evolutionary dynamics of cell and structure evolution in vertebrates. Notably, the origin and evolution of oligodendrocytes appears to be linked with specific genes that emerged within the gnathostome lineage. This suggests a deep association between the evolution of astrocytes, oligodendrocytes, and the evolutionary history of these species. Further analysis and data will help to unravel the complexities of these gene networks and their roles in glia evolution.

Additionally, focusing on the telencephalon has already provided important insights into the evolutionary history of inhibitory and excitatory neurons and their collective interplay with pallium evolution. Yet, this story is far from complete. While I have presented preliminary findings, the Kaessmann Lab's sustained efforts to generate additional data, including spatial information, will provide clearer picture. It is noteworthy that this endeavor is at the forefront of comparative research, incorporating representative species from nearly all major vertebrate lineages in a comprehensive analysis. This approach will elucidate the origins and diversification of cell types, structures, and molecular programs within the forebrain of vertebrates, increasing our understanding of their roles in the evolutionary history of this clade. Importantly, the results from this endeavor will substantially contribute to the refinement of pallium evolution models, which, until now, have predominantly relied on amniote data with sparse input from anamniotes.

5.1 Perspectives in cellular and organ evolution

Classifying cells into 'types' has long been a goal for cell biologists. However, a consensus on precisely what defines a cell type has not been universally established[72,109–111]. As can happen in research on species, although unanimous agreement is far from being reached, an operational concept is essential for various endeavors. For instance, the evolutionary concept of cell type, which addresses the mechanisms underlying cell evolution, provides a framework for interspecies comparisons[71,76]. This approach also permits the inclusion of comparative methods within a phylogenetic context, an area that has gained attention in recent years[68,77,78].

Yet, the field of cell evolution still contains much undiscovered territory. Research such as the study presented here is instrumental in advancing our understanding of the origins and diversification of cells, especially given its focus on one of the most complex organs in terms of cellular diversity and function. However, a challenge that I encountered is the relatively unexplored link between cell and organ evolution. Since evolutionary forces operate at multiple biological levels, a comprehensive framework that integrates the evolution of different cell types within the same organ, the divergence of the same cell type across different organs, and the evolution of the organ itself, is crucial. For instance, exploring these connections will help future studies on the origin and evolution of the amygdala and the pallium. While these structures may be homologous between species at the organ or cell type family level, many of their constituent cell types might not be, as suggested by the cellular diversity found in the tetrapod amygdala and the distinct interneurons in the lamprey pallium. Comparisons across these levels of organization offer an opportunity to explore the previously mentioned questions.

For that purpose, after a robust cell classification based on the transcriptome is established, it is crucial to delve deeper into understanding cell localization, environment, connectivity, and phenotype, and to determine how these factors correlate with the molecular program, thus providing insights into cell organization and function. There has been progress with the integration of spatial transcriptomics and genome editing techniques, but a consistent application of a broader range of methodologies remains scarce, particularly those that provide insights into cell structure and phenotype. While some attempts have been made[112], their scope is often limited to a small number of organisms or organs.

This endeavor requires collaboration across multiple disciplines and research groups. The integration of cell and organ specialists, who possess extensive knowledge and experience with specific tissues, is critical. Their expertise will greatly benefit the field of evolutionary biology, and conversely, their disciplines will benefit from the establishment of a robust, operational evolutionary framework for studying cells and organs. Such a framework will facilitate cross-species comparisons across various fields, aiding in the strategic selection of models for biomedical research as well as contributing to our understanding of cell fate specification and the mechanisms behind aberrant cell identity shifts. Just as evolutionary theory has contributed to the study of species since the days of the naturalists— and in turn has been enriched by their extensive data collection and knowledge— an integrative approach, akin to 'integrative taxonomy', will facilitate this endeavor [1,72].

Finally, while doing this, we must avoid the over-reliance on theoretical concepts without sufficient empirical data, as this could obscure the field. A thorough comprehension of the interplay between genes, cells, and organs, as well as their evolutionary trajectories, will illuminate the mechanisms by which evolution operates across various biological levels, much as it happens with populations, species, and ecosystems.

6 Methodology

This project represented the initial phase of a broader, ongoing research endeavor. As such, most activities were—and continue to be—conducted in association with fellow members of the Kaessmann lab and collaborators from other groups. The processing of animals and the collection of samples for hagfish, catshark, and lamprey embryos were carried out by multiple collaborators (described in section 6.1). I was responsible for euthanizing adult and larval lampreys, spotted gars, and lungfish. The dissections of lamprey (non-embryonic), spotted gar, and lungfish, as well as sample collection, data generation, and sequencing, were performed collaboratively with Dr. Phil Oel, Dr. Kerry Lynn Gendreau, Matthias Janeschik, and Dr. Xuefei Yuan, with the assistance of Céline Schneider and Julia Schmidt. Orthology assignment analyses were conducted in association with Dr. Francesco Lamanna and Matthias Janeschik. Dr. Francesco Lamanna also carried out the data processing and analyses for non-embryonic lampreys. All analyses for the other species were undertaken in collaboration with Dr. Kerry Lynn Gendreau, Matthias Janeschik, and Daniel Soto Carballo. Nevertheless, I created all the plots and performed all interspecies comparisons, for these last species, presented in this work. Additionally, I was responsible for all the data generation and analyses for lamprey embryonic stages, as well as the spatial transcriptomics approaches.

6.1 Animal handling and sample collection

We collected brain samples from five species: the sea lamprey (*Petromyzon marinus*), inshore hagfish (*Eptatretus burgeri*), small-spotted catshark (*Scyliorhinus canicula*), spotted gar (*Lepisosteus oculatus*), and West African lungfish (*Protopterus annectens*). For the sea lamprey, samples were taken at three life stages: adults, larvae (ammocoete), and embryos (Tahara stages 26 and 30). For the remaining species, only adult tissues were collected. The samples were obtained from various sources. We acquired adult sea lampreys, lungfish, and spotted gars from commercial suppliers in Spain (Novas Y Mar, Galicia; for lampreys) and Germany (Zierfischtreff, Theilenhofen). Some adult brains (used for spatial transcriptomics) were collected and embedded in OCT by Marianne Bronner and her collaborators, from the California Institute of Technology in Pasadena, USA. Larval sea lampreys were captured from the River Ulla in Galicia, Spain, by Antón Barreiro-Iglesias and his team (CIBUS, Universidade de Santiago de Compostela, Santiago de Compostela, Spain), shipped and maintained at Heidelberg University's Interfaculty Biomedical Research Facility under optimal

conditions until needed for tissue collection. Daniel M. Medeiros and colleagues at the University of Colorado, Boulder, retrieved and dissected the lamprey embryos. Hagfish were collected along the Korean coast, euthanized, and dissected by Tae Sung Jung and his group (College of Veterinary Medicine, Gyeongsang National University). French northern coast catsharks were euthanized and dissected by Sylvie Mazan and her team (CNRS, Station biologique de Roscoff).

I euthanized larval and adult lampreys, lungfish and spotted gars using a 0.1% MS-222 solution (Sigma, A5040-25G), adhering to local decapitation guidelines. We then air-dissected the brains, placed them in 1x HBSS (Life Technologies, 14185052) for cleaning, and meticulously removed the meninges. Cleaned brains were either kept whole or sectioned into regions for subsequent experiments. We processed samples based on the designated experiment, which included single-cell analysis, single-nuclei RNA-seq, or spatial transcriptomics. We performed all procedures with the animals following the ethical guidelines on animal care and experimentation of the European Union and Germany, with approval from the local animal welfare authorities (Regierungspräsidium) in Karlsruhe.

6.2 Genome annotation

In this section, I outline the procedure I employed for improving the annotation of the lungfish genome. I followed a workflow developed for other species by Dr. Kerry Lynn Gendreau. For the lamprey genome, the modifications implemented to generate a new annotation are described in Lamanna, Hervas-Sotomayor et al[5].

I used a total of thirteen libraries (ten publicly available[113]) for reannotating the lungfish genome as follows: The libraries were subjected to trimming using Trimmomatic[114] and verified with Fastqc (v0.11.9)[115]. After softmasking the published lungfish genome assembly[113] with RepeatMasker (v4.0)[116], I mapped the RNA reads to the genome using STAR v2.7[117]. The resulting BAM files were utilized to run BRAKER (v2.1.6)[118,119,128,120–127], a fully automated method that integrates GeneMark-ET[129] and AUGUSTUS[119,120]. BRAKER employs genomic and RNA-Seq data to automatically generate comprehensive gene structure annotations[121]. Inclusion of the mouse proteome to improve the prediction was performed by utilizing ProtHint mapping[129]. Due to the large size of the lungfish genome (43 Gb), I divided it in six parts, while preserving chromosome integrity. I executed the workflow above on each of these genome

segments. Once the BRAKER projection was completed, I concatenated the resulting six GTF files into a single file. Subsequent steps to generate a new annotation of the genome are currently underway in the Kaessmann lab by Daniel Soto Carballo. Therefore, I performed all the subsequent analyses by using the original annotation of the lungfish genome[113].

6.3 Orthology assignment

We identified gene correspondences between different organisms using OrthoFinder (v2.3.11)[130], a phylogeny-based orthology inference method. The OrthoFinder pipeline was run for specific chordate species, selected through unbiased phylogenetic sampling. These species include the European lancelet (*Branchiostoma lanceolatum*)[131], vase tunicate (*Ciona intestinalis*)[132], inshore hagfish (*Eptatretus burgeri*)[90], sea lamprey (*Petromyzon marinus*)[89], Elephant shark (*Callorhinchus milii*)[133], small-spotted catshark (*Scyliorhinus canicula*)[134], spotted gar (*Lepisosteus oculatus*)[135], zebrafish (*Danio rerio*)[136], West African lungfish (*Protopterus annectens*)[113], West Indian Ocean coelacanth (*Latimeria chalumnae*)[137], Western clawed frog (*Xenopus tropicalis*)[138], red junglefowl (*Gallus gallus*)[139], house mouse (*Mus musculus,* GRCm38)[140], and human (*Homo sapiens*)[140].

We reconstructed a comprehensive set of gene trees with root nodes for the analyzed species. This facilitated the establishment of orthology relationships among all genes, inference of duplication events, and cross-referencing them to corresponding nodes on the gene and species trees. We only used peptides derived from the longest isoform within each gene. Rooted gene trees were derived from orthogroups (groups of genes descended from a single gene in the Last Chordate Ancestor) using Multiple Sequence Alignments (MAFFT v7.455[141]) with IQ-TREE v1.6.12[142] (1,000 bootstrap replicates) and STRIDE[143].

Orthology relationships for a subset of these species can be explored in the online atlas from Lamanna, Hervas-Sotomayor et al.[5] (lampreybrain.kaessmannlab.org). There is not a consensus for a gene name format in vertebrates, so given that most of the species I analyzed here are gnathostomes I don't capitalize the gene names as we did for the lamprey genes in Lamanna, Hervas-Sotomayor et al.[5]. I instead used mouse orthologous gene names for all the species, following Woych et al.[48]. In cases where multiple mouse genes correspond to one gene in the considered species (one-to-many relationships), I appended all orthologues names, separated by slashes.

6.4 Tissue dissociation and sn/scRNA-seq data generation

Dr. Phil Oel and I conducted cell dissociation and single-cell RNA-sequencing protocols on larval and adult lamprey brains, as described in Lamanna, Hervas-Sotomayor et al[5]. In summary, fresh whole brains and brain parts were dissociated using the Papain Dissociation System (Worthington, LK003150) with minor modifications to the manufacturer's protocol. Initially, the tissue was incubated in a papain solution at a constant agitation of 28 ºC for 15 minutes. Subsequently, the tissue was gently triturated by pipetting up and down, and the dissociated cells were collected by centrifugation at 300g for 1 minute. This process was followed by a second incubation in fresh papain solution and a final trituration step, similar to the previous one. The dissociated cells were then centrifuged at 300g for 5 minutes and resuspended in an inhibitor solution, prepared according to the specifications of the Papain Dissociation System. To remove larger cell aggregates, the suspension was filtered through a 40 µm falcon strainer (Sigma-Aldrich, CLS431750-50EA), and immediately after, a discontinuous density gradient was performed. Subsequently, the cells were resuspended in Leibovitz's L-15 Medium (Life Technologies, 21083027) to achieve a final volume of 50 to 100 µl, depending on the original tissue size. Cell viability and concentration were assessed by trypan blue staining and counting using a Neubauer counting chamber (Assistent). To ensure high viability, cells with a viability greater than 90% were selected, and the cell concentration was equal to or higher than 300 cells per µl.

For the adult brains from other species, Dr. Kerry Lynn Gendreau, Matthias Janeschik, Dr. Xuefei Yuan and myself performed nuclei dissociation and single nucleus RNA-sequencing on previously snap-frozen telencephaloni from hagfish, catshark, spotted gar and lungfish, following Sarropoulos et al.[144], with some modifications. I followed the same protocol for dissociating lamprey embryonic heads. Briefly, while working on ice, we incubated the frozen brains for 20 seconds in a solution containing 250 mM sucrose, 25 mM KCl, 5 mM MgCl2, 10 mM Tris-HCl (pH 8), 0.1% IGEPAL, 1 µM DTT, 0.4 U/µl Murine RNase Inhibitor (New England BioLabs), and 0.2 U/µl SUPERas-In (Ambion). After incubation, tissues were homogenized by pipette trituration and using a micropestle. After 5 minutes of incubation, the remaining unlysed tissue was pelleted by centrifugation at 100 g for 1 minute. We then centrifuged the cleared homogenate at 400 g for 4 minutes to pellet the nuclei. Nuclei were washed in the homogenization buffer once, collected by centrifugation, and resuspended in filtered 1X PBS. The nuclei were strained using 40 µM Falcon (Sigma-Aldrich, CLS431750-

50EA) or Flowmi strainers (Sigma), depending on the final resuspension volume. We estimated the nuclei concentration on a Countess II FL Automated Cell Counter (Thermo Fisher Scientific) by performing a double staining with Hoechst DNA dye and PI.

We loaded cell and nuclei suspensions (approximately 15,000 cells/nuclei per reaction) onto the Chromium system (10x Genomics). cDNA was amplified, and RNA-seq libraries were constructed using Single-Cell 3′ Gel Bead and Library v2 kits (for lamprey larvae), v3 kits (for all other animal brains and a few lamprey larvae), and Multiome ATAC + Gene Expression kits (for lamprey embryos), following the manufacturer's instructions. The cDNA libraries were amplified using 12-13 PCR cycles and quantified using a Qubit Fluorometer (Thermo Fisher Scientific). The average fragment size was determined using a Fragment Analyzer (Agilent). The libraries were sequenced using the NextSeq 500/550 High Output Kit v2.5 on the Illumina NextSeq 550 system, with 28 cycles for Read 1, 56 cycles for Read 2, 8 cycles for i7 index, and 0 cycles for i5 index. For each library, we allocated approximately half a lane for sequencing, resulting in approximately 20,000 reads per cell/nuclei.

6.5 Single cell and single nuclei RNA data processing

The analysis of adult and larvae lamprey single-cell RNA-seq data is described in Lamanna, Hervas-Sotomayor et al.[5]. I performed the sequence mapping, quality control, clustering, integration and broad cell annotation for lamprey embryonic and lungfish data, as described below. A similar procedure was done for spotted gar and catshark data, by Dr. Kerry Lynn Gendreau and Matthias Janeschik, respectively. I aligned the single-nucleus RNA-seq reads to their respective reference genomes, and generated unique molecule identifier (UMI) count matrices using CellRanger ARC (10x Genomics), for lamprey embryonic data, and STARsolo v2.7[117], lungfish.

The identification of droplets containing cells/nuclei was obtained through the CellRanger and STARsolo calling algorithms, but I validated these droplets by assessing the following criteria:

1. The cumulative distribution of UMIs.
2. The distribution of UMIs derived from mitochondrial genes.
3. The distribution of the proportion of UMIs originating from intronic regions.

Empty droplets containing ambient RNA were identified either through DropletUtils[145]or by visually inspecting the cumulative curves. I detected putative multiplets, which are droplets

containing more than one cell/nucleus, using scDblFinder v3.16[146] and subsequently excluded them from the count matrices.

I analyzed the obtained count matrices using Seurat v4[147]. Pre-processing steps involved retaining only genes expressed in at least five nuclei and removing nuclei that contained less than 200 UMIs or more than 5% mitochondrial UMIs. Next, I normalized raw UMI counts using the SCTransform method[148]. I selected the top 3,000 Highly Variable Genes (HVGs) across all nuclei for subsequent clustering, and applied Principal Component Analysis (PCA) to the normalized HVG matrices. The 50 most significant principal components (PCs) were used to construct a Shared Nearest Neighbor (SNN) graph. The SNN graph was then clustered using the Louvain method with various resolution values ranging from 0.5 to 10. I performed differential expression analysis to identify potential marker genes from all clusters at each resolution value, using the Wilcoxon Rank Sum Test with the criteria logFC $\geq$ 0.25, min.pct = 0.1, and Bonferroni-adjusted p-value < 0.01. The PCA-transformed matrices were further embedded into a two-dimensional space using the Uniform Manifold Approximation and Projection (UMAP) dimensionality reduction technique. To ensure data quality, I inspected the clustered cells to remove any spurious clusters. Notably, for the lamprey embryos, I only analyzed the transcriptome dataset (ATAC-seq data was also generated by using the multiome kit).

6.6 Cell annotation

Cell class/type annotation performed for the adult and larval lamprey datasets is described in Lamanna, Hervas-Sotomayor et al[5]. Following a consistent approach, I performed a broad cell class annotation for lamprey embryonic and lungfish data. Similarity, Dr. Kerry Lynn Gendreau and Matthias Janeschik annotated broad cell class in the spotted gar and catshark data, respectively. In brief, classes/types were annotated by performing hierarchical clustering and annotating higher resolution clusters first. A putative phenotype was assigned to each cluster by allocating marker genes to any of the following Gene Ontology[149] (GO) categories: transcription (co-)factor, neurotransmitter metabolism, neurotransmitter transport, neurotransmitter receptor, neuropeptide, neuropeptide receptor, immune response, erythrocyte differentiation, blood vessel development, neurogenesis, gliogenesis. Additional functional information was added by comparing the annotated clusters to published vertebrate neural

single-cell datasets. The annotation of cluster a higher resolution is an ongoing effort for lungfish, catshark and spotted gar.

6.7 Data integration

Dr. Francesco Lamanna integrated all adult and larvae lamprey datasets coming from different samples using integrative non-Negative Matrix Factorization (iNMF) as implemented in LIGER v0.5.0[150]. I integrated the lamprey embryonic data following the same procedure. For the other species, we (with Dr. Kerry Lynn Gendreau and Matthias Janeschik) accessed the performance of two integration methods: LIGER v0.5.0[150] and Harmony[151], an algorithm that projects cells into a shared embedding in which cells group by cell type rather than dataset-specific conditions. Both algorithms yielded similar results, but because Harmony requires less computing time, we decided to utilize it as follows: matrices coming from the same species were filtered and then individually SCTransformed[148]. We then merged and integrated them. In each integrated dataset, we performed SNN graph construction, clustering, DE analysis, 2D-embedding and cluster annotation as described above.

6.8 Species comparisons

6.8.1 Lamprey-mouse comparisons

In order to find cross-vertebrate similarities and differences in neural cell types, the lamprey adult brain atlas was compared against a published juvenile mouse nervous system atlas[84]. These analyses were done by Dr. Francesco Lamanna. In summary, he compared the two datasets via a correlation-based approach. That is, the raw UMI count matrices were extracted from both species datasets and orthology information for the corresponding gene IDs was added; orthology relationships between mouse and lamprey were obtained from the OrthoFinder analysis (see above). The UMI counts coming from paralogues in the respective species were summed, using the "Meta-gene" approach[79], and the species-specific gene IDs were replaced by numeric indexes (1..*n*, where *n* is the number of the orthology group between the mouse and lamprey) shared by the two species. The new "meta-gene" count matrices were then normalized using SCTransform, filtered for HVGs, and averaged across all annotated clusters. Expression levels were finally transformed to gene specificity indexes (SI), calculated following Tosches et al[52]. To obtain the SIs with this method, the mean of normalized scRNA-

seq read counts of each gene (g_c) was calculated for each cell type (C) and then divided by its mean across all cells:

$$s_{g,c} = \frac{g_c}{\frac{1}{N}\sum_{i \in c} g_i}$$

Finally, he used the SIs for Pearson correlation analyses and constructed dendrograms relating cell-type families between lamprey and mouse with the pvclust[152] R package, with complete hierarchical clustering and 1,000 replicates.

Additionally, he also compared the two datasets using the Self Assembling Manifold mapping (SAMap; v0.2.3) algorithm[77], a method that enables mapping single-cell transcriptomic atlases between phylogenetically distant species. A gene-gene bipartite graph with cross species edges connecting homologous gene pairs was constructed by performing reciprocal BlastP searches between the two proteomes of the two species. The graph was used in a second step to project the two datasets into a joint, lower-dimensional manifold representation, where expression correlation between homologous genes was iteratively used to update the homology graph connecting the two atlases. After the analysis was run, a mapping score (ranging from 0 to 1) was computed among all possible cross-species cluster pairs.

6.8.2 Comparisons of the pallium in gnathostomes.

I performed the correlations among pallial cells of gnathostomes (catshark, spotted gar, lungfish), and included published data for the Spanish ribbed newt (*Pleurodeles waltl*)[48]. I followed the same procedure previously described for the lamprey-mouse correlation, employing the "Meta-gene" method[79]. For these analyses, I generated two types of similarity index (SI) matrices: one derived exclusively from filtering for transcription factors (137 orthogroups), using the Human Transcription Factors database[153], and the other encompassing the entire transcriptome (4,546 orthogroups). These matrices were then used for Spearman correlation analyses and the construction of dendrograms as previously detailed. I repeated the analysis on a subset of the data by downsampling the number of cells per cluster to 500 to evaluate the influence of varying cluster sizes on the correlation, although few clusters maintained smaller cell numbers even after the downsampling (no less than 200).

6.9 Spatial transcriptomics

All spatial transcriptomics assays, as well as the whole microscopy, showed here and in Lamanna, Hervas-Sotomayor et al.[5] were performed by me. However, some of these protocols were adapted from the previous optimizations made by Dr. Mari Sepp, a Postdoctoral researcher in the Kaessmann Lab.

6.9.1 In situ sequencing

The brains of adult lampreys (performed by our collaborators, as described in section 6.1) and whole heads of larvae were embedded in OCT mounting medium and rapidly frozen by placing them on isopentane, which had been pre-cooled in liquid nitrogen. I later continued with the cryosectioning of the frozen tissues, producing 10 µm thick coronal and sagittal sections that were then stored at -80 °C until further use. For *in situ* sequencing, I processed the sections using the High Sensitivity Library Preparation Kit from CARTANA AB (10x Genomics), following the method and data processing described by Ke et al.[154], with minor modifications. Briefly, the sections were air dried for 5 minutes on SuperFrost Plus glass slides (Thermo Fisher Scientific). Subsequently, I fixed the sections with 3.7% paraformaldehyde in UltraPure distilled water (DNase/RNase-Free, Thermo Fisher Scientific, 10977035) for 7 minutes, continuing with washing with PBS (diluted in UltraPure distilled water, Thermo Fisher Scientific, 70011036). Next, the sections underwent a 5-minute treatment with 0.1 N HCl and were washed again with PBS. Ethanol dehydration was then carried out, followed by air drying. The sections were covered with SecureSeal hybridization chambers (Grace Bio-Labs, 10910000). For all subsequent steps, including probe hybridization and ligation, amplification, fluorescent labeling, and quality control imaging, I followed the specifications provided by the manufacturer. Finally, I shipped the mounted sections to CARTANA's facility in Solna, Sweden, for *in situ* sequencing.

6.9.2 Single molecule RNA-FISH

The whole heads of lamprey larvae were snap-frozen and cryosectioned horizontally, as mentioned previously. However, in this case, I collected the sections on coverslips (22 mm x 22 mm) that had been pretreated with a silanization solution (0.3% bind-silane, 0.1% acetic acid, and 99.6% ethanol, GE Healthcare Life Sciences, 17-1330-01). To minimize tissue

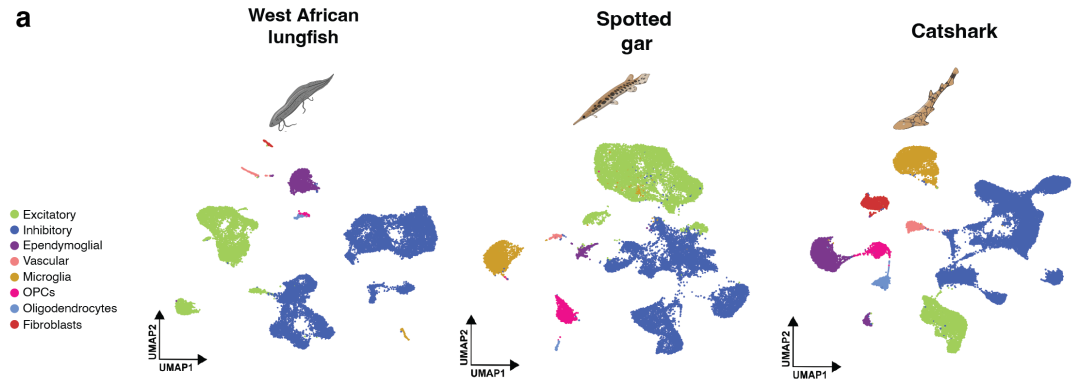
autofluorescence, I embedded the sections in a polyacrylamide gel where RNAs were anchored using LabelX treatment, while cellular proteins and lipids were cleared. LabelX solution was prepared by combining Label-IT (Mirus Bio) with Acryloyl X - SE (Thermo Fisher Scientific), following Chen et al.[155]. Specifically, I air-dried the sections for 15-20 minutes and fixed them in 3.7% paraformaldehyde in PBS for 10-15 minutes, followed by a 2-minute incubation in 4% SDS in PBS and subsequent PBS washes. The fixed sections were then incubated in 70% ethanol at 4°C for at least 16 hours.

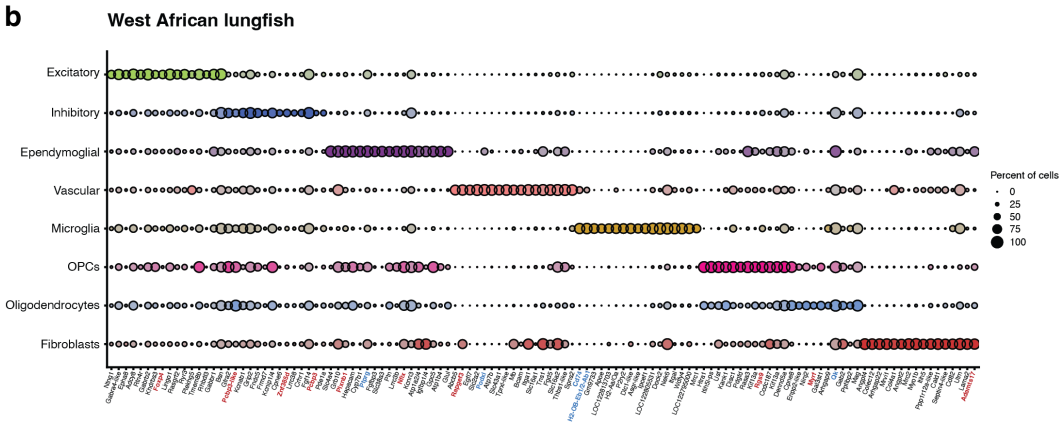
Next, I washed the sections with PBS and a wash with 1x MOPS pH 7.7 (Sigma-Aldrich, M9381), and subsequently I incubated them with LabelX diluted in 1x MOPS (at a concentration of 0.006 mg/mL) for 4 hours at room temperature. After two PBS washes, the LabelX-modified RNAs were anchored by embedding the sections in thin 4% polyacrylamide (PA) gels. The coverslips were washed with a PA solution containing 4% (v/v) 19:1 acrylamide/bis-acrylamide, 60 mM Tris·HCl pH 8, and 0.3 M NaCl. Then, the coverslips were washed with the same PA solution supplemented with ammonium persulfate and TEMED. To cast the gel, a drop of the PA solution (containing the polymerizing agents) was added to Repel Silane-treated glass slides and covered with each coverslip, ensuring the formation of a thin PA layer between the slide and the coverslip. The gel was allowed to solidify at room temperature for 1.5 hours. Coverslips and slides were gently separated, leaving the coverslips with the sections embedded in the PA gel. Subsequently, the coverslips were washed with a digestion buffer consisting of 0.8 M guanidine-HCl, 50 mM Tris·HCl pH 8, 1 mM EDTA, and 0.5% (v/v) Triton X-100. Digestion buffer supplemented with proteinase K (8 U/ml) was used for incubation at 37°C for 2-3 hours to digest the gel.

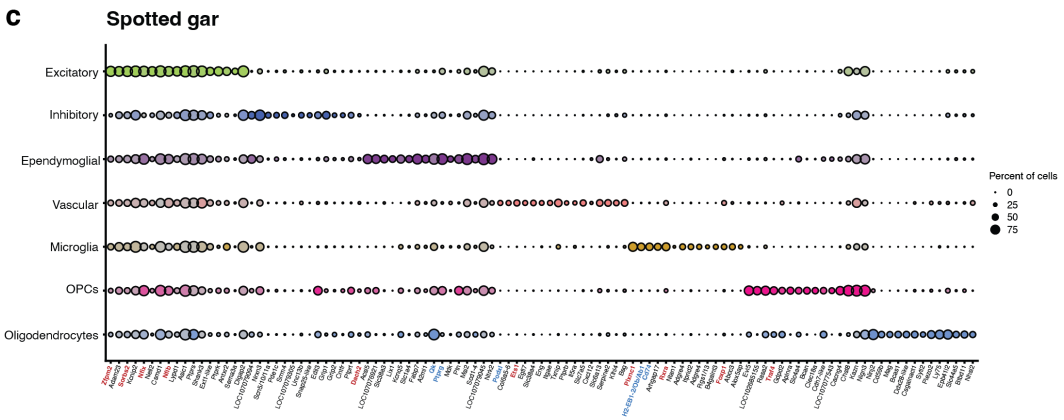
Following background reduction, I continue with the hybridization with HuluFISH probes developed by PixelBiotech. The hybridization protocol followed the manufacturer's recommendations. Coverslips were washed with HuluWash buffer (PixelBiotech GmbH) and incubated in a probe solution containing each probe at a concentration of 1:100 diluted in hybridization buffer. The coverslips were incubated at 37°C for 12 hours in a light-protected humidified chamber. After four washes with HuluWash buffer (each lasting 10 minutes at room temperature), the final wash was supplemented with Hoechst 33342. Finally, the coverslips were mounted in Prolong Diamond mounting medium and allowed to cure for 24 hours at room temperature.

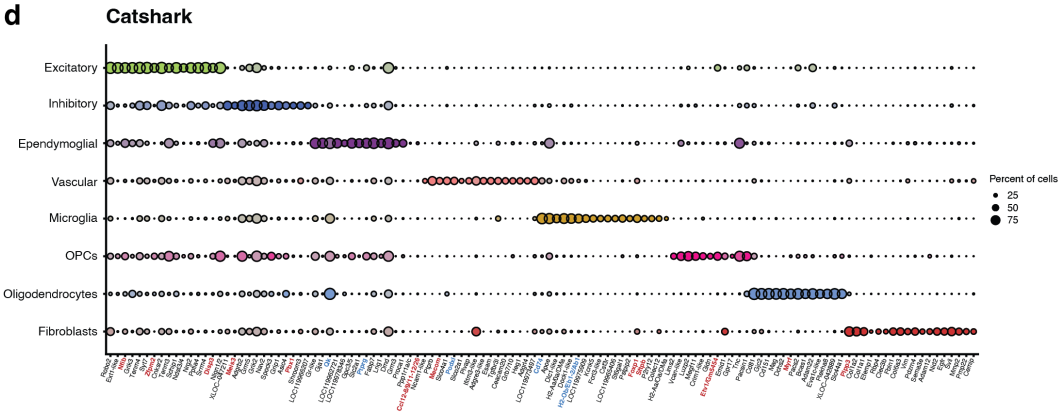
I imaged all sections using a Leica TCS-SP5 confocal laser scanning microscope controlled by the Leica Application Suite (LAS). The displayed images are projections of mosaics created by stitching individual z-stacks. Each z-stack comprised multiple images (50 images for *Sspoa*,) taken within a range of 10-15 µm, with a step size below 0.8 µm. Imaging was performed using a 63x immersion oil objective, sequentially exciting the samples with a 405 nm Diode laser (for Hoechst 33342 staining), followed by the appropriate laser for each probe (561 nm DPSS laser for *Sspoa*). I generated projections of the z-stacks using Fiji 2109, applying average intensity projection. Additional processing, when necessary, included contrast enhancement (saturated pixels between 0.1 and 0.3%) and background subtraction for noise reduction using a rolling ball algorithm with a radius of 50 pixels.

7 Supplementary Figures

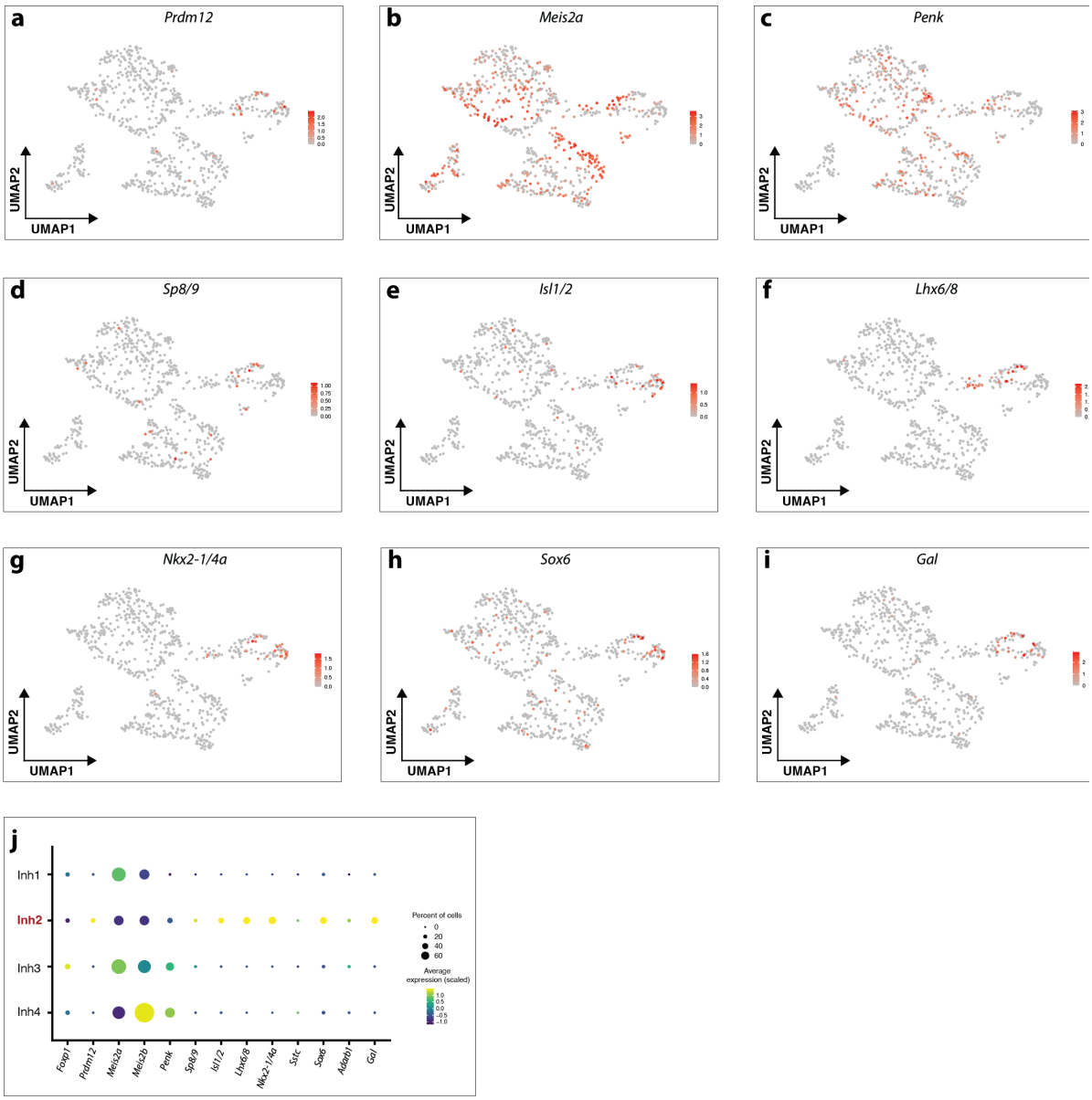




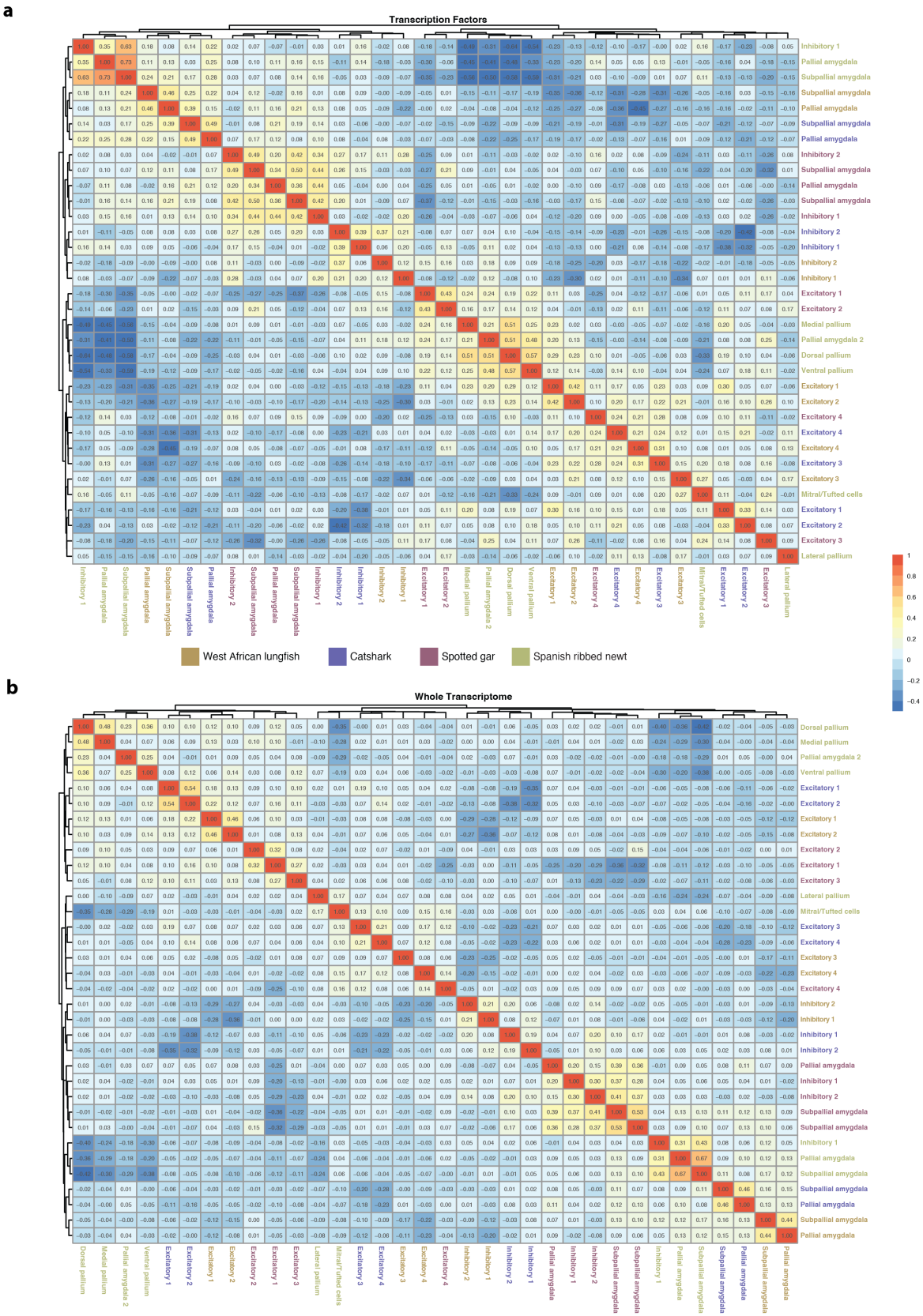




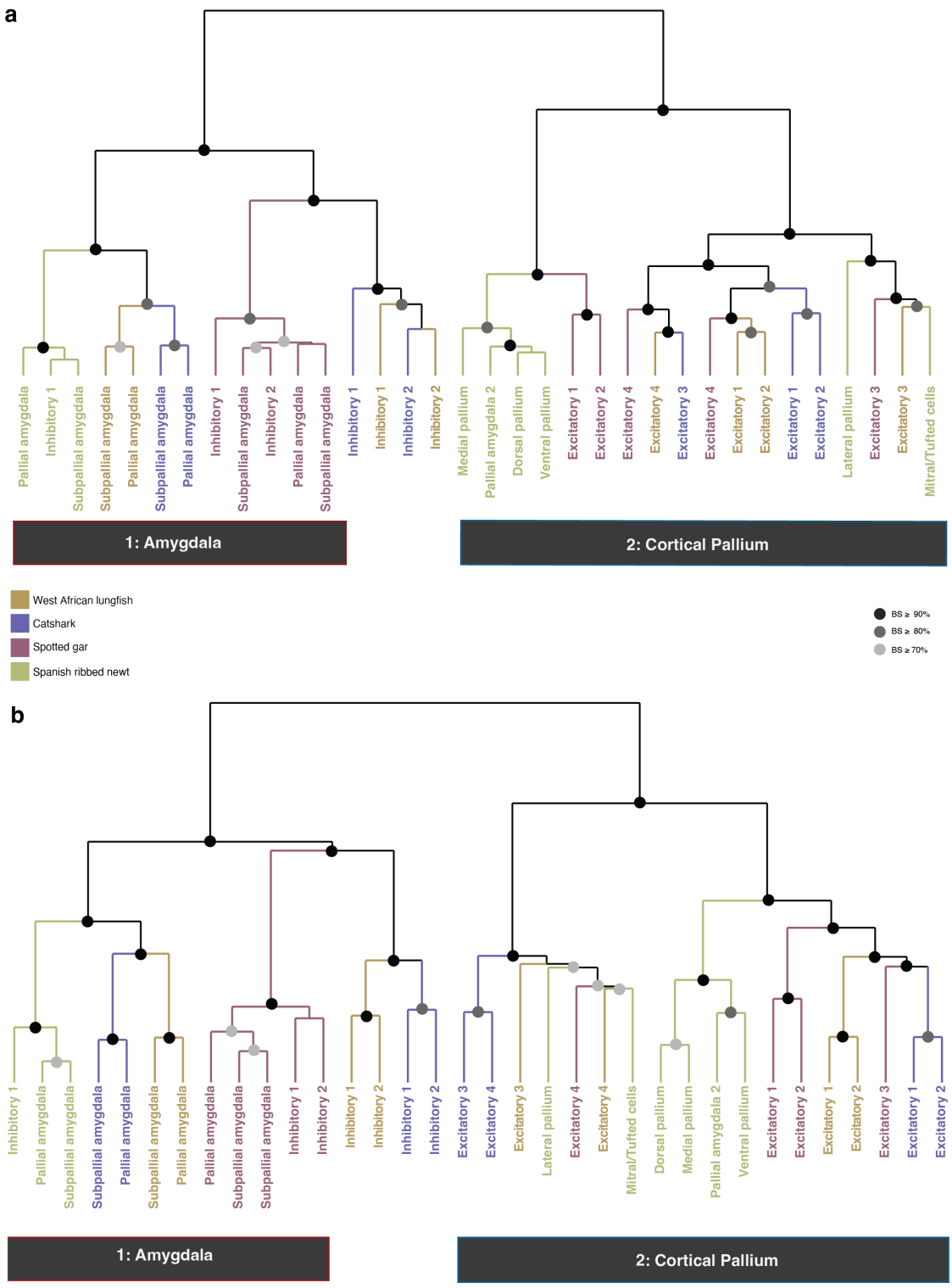
Supplementary Figure 1. Lungfish, Spotted gar, and Catshark brain cell classes and their molecular specification. a) UMAP representation of cells (combining all single-cell RNA sequencing data) colored by their respective cell type classes. b-c) Expression of transcription factors (in red) and effector genes within each cell class. Genes from Figs. 6-7 are excluded here. Genes shared among these three species are highlighted in blue.



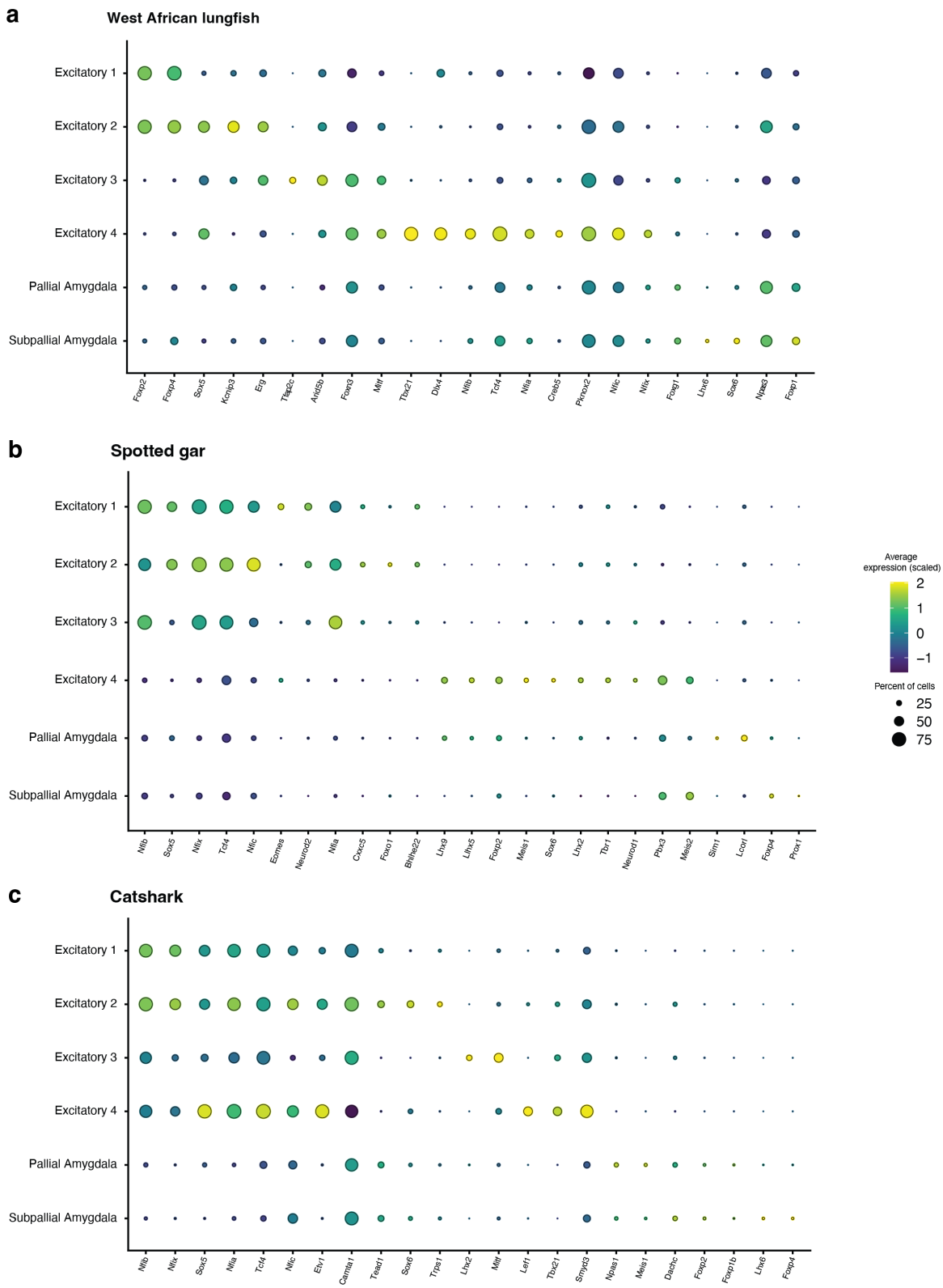
Supplementary Figure 2. Gene expression in developmental sub-pallial cells. a-i) Feature plots display the expression of genes associated with the medial ganglionic eminence (MGE) and lateral ganglionic eminence (LGE). j) Dot plot depicting that one out of four inhibitory neuron types express the genes shown in a-i. The data correspond to embryos at Tahara stage 26.



Supplementary Figure 3. Correlations between pallial cells across gnathostomes. Heatmaps present Spearman's ρ (rank correlation coefficient) of specificity indices for cell groups for transcription factors only (a) and for all orthologous genes (b). These analyses are based on orthogroups following the "Meta-gene" approach.



Supplementary Figure 4. Comparison of telencephalic excitatory neurons in gnathostomes (downsampled data). Dendrograms illustrate the gene expression distance (Spearman's ρ) for a) transcription factors (136 orthogroups) and b) the whole transcriptome (4,546 orthogroups). The data were downsampled to a uniform cluster size of 500 cells, although some clusters retained fewer cells (no fewer than 200). Different colors represent distinct species. Approximately unbiased (au) bootstrap support (BS) is provided (n = 1,000).



Supplementary Figure 5. Expression profiles of transcription factors in gnathostome pallial cells. Transcription factors selected from the top 200 differentially expressed genes between cell classes for lungfish (a), spotted gar (b), and catshark (c). Common genes between species are depicted comparatively in Fig. 29b.

8 Literature cited

1. Striedter, G. & Northcutt, R. G. *Brains Through Time: A Natural History of Vertebrates*. (NY: Oxford University Press., 2020).
2. Butler, A. & Hodos, W. *Comparative Vertebrate Neuroanatomy*. (2005).
3. Sugahara, F., Murakami, Y., Pascual-Anaya, J. & Kuratani, S. Reconstructing the ancestral vertebrate brain. *Dev. Growth Differ.* **59**, 163–174 (2017).
4. Northcutt, R. G. Understanding vertebrate brain evolution. *Integr. Comp. Biol.* **42**, 743–756 (2002).
5. Lamanna, F. *et al.* A lamprey neural cell type atlas illuminates the origins of the vertebrate brain. *Nat. Ecol. Evol.* **7**, 1714–1728 (2023).
6. Sugahara, F., Murakami, Y., Pascual-Anaya, J. & Kuratani, S. Reconstructing the ancestral vertebrate brain. *Dev. Growth Differ.* **59**, 163–174 (2017).
7. Wicht, H. The brains of lampreys and hagfishes: characteristics, characters and comparisons. *Brain, Behav. Evol.* **48**, 248 – 261 (1996).
8. Robertson, B. *et al.* The lamprey blueprint of the mammalian nervous system. *Prog. Brain Res.* **212**, 337–349 (2014).
9. Pani, A. M. *et al.* Ancient deuterostome origins of vertebrate brain signalling centres. *Nature* **483**, 289–294 (2012).
10. Vopalensky, P. *et al.* Molecular analysis of the amphioxus frontal eye unravels the evolutionary origin of the retina and pigment cells of the vertebrate eye. *Proc. Natl Acad. Sci. U. S. A* **109**, 15383–15388 (2012).
11. Murakami, Y., Uchida, K., Rijli, F. M. & Kuratani, S. Evolution of the brain developmental plan: Insights from agnathans. *Dev. Biol.* **280**, 249–259 (2005).
12. Williams, N. A. & Holland, P. W. H. Molecular Evolution of the Brain of Chordates. *Brain. Behav. Evol.* **52**, 177–185 (1998).
13. Gilbert, S. . *Developmental Biology*. (Sinauer Associates, Incorporated Publishers, 2014).
14. Rubenstein, J. L. R., Martinez, S., Shimamura, K. & Nina, L. P. The embryonic vertebrate forebrain: The prosomeric model. *Science (80-.).* **266**, 578–580 (1994).
15. Puelles, L., Harrison, M., Paxinos, G. & Watson, C. A developmental ontology for the mammalian brain based on the prosomeric model. *Trends Neurosci.* **36**, 570–578 (2013).
16. Nieuwenhuys, R. & Nicholson, C. The Central Nervous System Vertebrates. in (ed.

Nieuwenhuys, R. et al) 397–495 (Springer, 1998).

17. Holland, L. Z. *et al.* Evolution of bilaterian central nervous systems: A single origin? *Evodevo* **4**, 1–20 (2013).
18. Benito-Gutiérrez, È. *et al.* The dorsoanterior brain of adult amphioxus shares similarities in expression profile and neuronal composition with the vertebrate telencephalon. *BMC Biol.* **19**, 1–19 (2021).
19. Cao, C. *et al.* Comprehensive single-cell transcriptome lineages of a proto-vertebrate. *Nature* **571**, 349–354 (2019).
20. Albuixech-Crespo, B. *et al.* Molecular regionalization of the developing amphioxus neural tube challenges major partitions of the vertebrate brain. *PLoS Biology* **15**, (2017).
21. Gans, C. & Northcutt, R. Neural crest and the origin of vertebrates: a new head. *Science (80-.).* **220**, 268–273 (1983).
22. Wullimann, M., Rupp, B. & Reichert, H. *Neuroanatomy of the Zebrafish Brain: A Topological Atlas*. (Elsevier Science Ltd. Published by Elsevier Inc., 1996). doi:: https://doi.org/10.1016/S0166-2236(96)60028-9
23. Puelles, L. *et al. The Pallium in Reptiles and Birds in the Light of the Updated Tetrapartite Pallium Model*. *Evolution of Nervous Systems: Second Edition* **1**–**4**, (2016).
24. Grillner, S. & Robertson, B. The basal ganglia downstream control of brainstem motor centres-an evolutionarily conserved strategy. *Curr. Opin. Neurobiol.* **33**, 47–52 (2015).
25. Smulders, T. V. Opinion piece. The relevance of brain evolution for the biomedical sciences. *Biol. Lett.* **5**, 138–140 (2009).
26. Krubitzer, L. In search of a unifying theory of complex brain evolution. *Ann N Y Acad Sci.* **1156**, 44–67 (2009).
27. Striedter, G. F. & Northcutt, R. G. The Independent Evolution of Dorsal Pallia in Multiple Vertebrate Lineages. *Brain. Behav. Evol.* **96**, 200–211 (2022).
28. Kumar, S., Stecher, G., Suleski, M. & Hedges, S. TimeTree: a resource for timelines, timetrees, and divergence times. *Mol. Biol. Evol.* **34**, 1812–1819 (2017).
29. Grillner, S., von Twickel, A. & Robertson, B. The blueprint of the vertebrate forebrain – With special reference to the habenulae. *Semin. Cell Dev. Biol.* **78**, 103–106 (2018).
30. Osório, J. & Rétaux, S. The lamprey in evolutionary studies. *Dev. Genes Evol.* **218**, 221–235 (2008).
31. Barreiro-Iglesias, A. *et al.* The sea lamprey tyrosine hydroxylase: cDNA cloning and in Situ hybridization study in the brain. *Neuroscience* **168**, 659–669 (2010).
32. Kebschull, J. M. *et al. Cerebellum Lecture: the Cerebellar Nuclei—Core of the*

Cerebellum. *Cerebellum* (Springer US, 2023). doi:10.1007/s12311-022-01506-0

33. Sugahara, F. *et al.* Evidence from cyclostomes for complex regionalization of the ancestral vertebrate brain. *Nature* **531**, 97–100 (2016).
34. Kuratani, S., Kuraku, S. & Murakami, Y. Lamprey as an evo-devo model: Lessons from comparative embryology and molecular phylogenetics. *Genesis* **34**, 175–183 (2002).
35. Stesiö, E. *The Downtonian and Devonian Vertebrates of Spitsbergen: Part 1. Family Cephalaspidae*. (I kommisjon hos J. Dybwad, 1927).
36. Briscoe, S. D. & Ragsdale, C. W. Homology, neocortex, and the evolution of developmental mechanisms. *Science (80-.).* **362**, 190–193 (2018).
37. Sugahara, F. *et al.* Evidence from cyclostomes for complex regionalization of the ancestral vertebrate brain. *Nature* **531**, 97–100 (2016).
38. Ericsson, J., Silberberg, G., Robertson, B., Wikström, M. A. & Grillner, S. Striatal cellular properties conserved from lampreys to mammals. *J. Physiol.* **589**, 2979–2992 (2011).
39. Stephenson-Jones, M., Samuelsson, E., Ericsson, J., Robertson, B. & Grillner, S. Evolutionary conservation of the basal ganglia as a common vertebrate mechanism for action selection. *Curr. Biol.* **21**, 1081–1091 (2011).
40. Martínez-de-la-Torre, M., Pombal, M. A. & Puelles, L. Distal-less-like protein distribution in the larval lamprey forebrain. *Neuroscience* **178**, 270–284 (2011).
41. Pombal, M. A., Álvarez-Otero, R., Pérez-Fernández, J., Solveira, C. & Megías, M. Development and organization of the lamprey telencephalon with special reference to the GABAergic system. *Front. Neuroanat.* **5**, 3 (2011).
42. Yuan, T., York, J. R. & McCauley, D. W. Gliogenesis in lampreys shares gene regulatory interactions with oligodendrocyte development in jawed vertebrates. *Dev. Biol.* **441**, 176–190 (2018).
43. Smith, J. J. *et al.* The sea lamprey germline genome provides insights into programmed genome rearrangement and vertebrate evolution. *Nat. Genet.* **50**, 270–277 (2018).
44. Briscoe, S. D. & Ragsdale, C. W. Evolution of the Chordate Telencephalon. *Curr. Biol.* **29**, R647–R662 (2019).
45. Nieuwenhuys, R. & Puelles, L. *Towards a New Neuromorphology*. (Springer, 2016).
46. Puelles, L. Thoughts on the development, structure and evolution of the mammalian and avian telencephalic pallium. *Philos. Trans. R. Soc. B Biol. Sci.* **356**, 1583–1598 (2001).
47. Medina, L., Abellán, A. & Desfilis, E. Evolving Views on the Pallium. *Brain. Behav. Evol.* **96**, 181–199 (2022).

48. Woych, J. *et al.* Cell-type profiling in salamanders identifies innovations in vertebrate forebrain evolution. *Science (80-.).* **377**, (2022).
49. Northcutt, R. G. Understanding vertebrate brain evolution. *Integr. Comp. Biol.* **42**, 743–756 (2002).
50. Passingham, R. E. & Wise, S. P. *The Neurobiology of the Prefrontal Cortex: Anatomy, Evolution, and the Origin of Insight*. (Oxford University Press, 2012).
51. Porter, B. A. & Mueller, T. The Zebrafish Amygdaloid Complex – Functional Ground Plan, Molecular Delineation, and Everted Topology. *Front. Neurosci.* **14**, (2020).
52. Tosches, M. A. *et al.* Evolution of pallium, hippocampus, and cortical cell types revealed by single-cell transcriptomics in reptiles. *Science (80-.).* **360**, 881–888 (2018).
53. Medina, L. & Abellán, A. Development and evolution of the pallium. *Semin. Cell Dev. Biol.* **20**, 698–711 (2009).
54. Huesa, G., Anadón, R., Folgueira, M. & Yáñez, J. Evolution of the Pallium in Fishes BT - Encyclopedia of Neuroscience. in (eds. Binder, M. D., Hirokawa, N. & Windhorst, U.) 1400–1404 (Springer Berlin Heidelberg, 2009). doi:10.1007/978-3-540-29678-2_3166
55. Northcutt, R. G. & Wicht, H. Afferent and efferent connections of the lateral and medial pallia of the silver lamprey. *Brain. Behav. Evol.* **49**, 1–19 (1997).
56. Northcutt, R. G. & Puzdrowski, R. Projections of the olfactory bulb and nervus terminalis in the silver lamprey. *Brain. Behav. Evol.* **32**, 96–107 (1988).
57. Suryanarayana, S. M., Robertson, B., Wallén, P. & Grillner, S. The Lamprey Pallium Provides a Blueprint of the Mammalian Layered Cortex. *Curr. Biol.* **27**, 3264-3277.e5 (2017).
58. Ocaña, F. M. *et al.* The lamprey pallium provides a blueprint of the mammalian motor projections from cortex. *Curr. Biol.* **25**, 413–423 (2015).
59. Ebbesson, S. & L, H. Projections of the olfactory tract fibers in the nurse shark (*Ginglymostoma cirratum*). *Brain Res.* **17**, 47– 55 (1970).
60. Hofmann, M. H. & Northcutt, R. G. Forebrain organization in elasmobranchs. *Brain. Behav. Evol.* **80**, 142–151 (2012).
61. Murakami, T., Morita, Y. & Ito, H. Extrinsic and intrinsic fiber connections of the telencephalon in a teleost, *Sebastiscus marmoratus*. *J. Comp. Neurol.* **216**, 115–131 (1983).
62. Kanwal, J., Finger, T. & Caprio, J. Forebrain connections of the gustatory system in ictalurid catfishes. *J. Comp. Neurol.* **278**, 353–376 (1988).

63. Saidel, W., Marquez-Houston, K. & Butler, A. Identification of visual pallial telencephalon in the goldfish, *Carassius auratus*: a combined cytochrome oxidase and electrophysiological study. *Brain Res.* **919**, 82–93 (2001).
64. Nieuwenhuys, R. The forebrain of actinopterygians revisited. *Brain. Behav. Evol.* **73**, 229–252 (2009).
65. Northcutt, R. & Davis, R. Telencephalic organization in ray-finned fishes. in *Fish Neurobiology* (eds. Northcutt, R. & Davis, R.) 203–236 (U Michigan Press, 1983).
66. Pombal, M. A., Megías, M., Bardet, S. M. & Puelles, L. New and old thoughts on the segmental organization of the forebrain in lampreys. *Brain. Behav. Evol.* **74**, 7–19 (2009).
67. Osório, J., Megías, M., Pombal, M. A. & Rétaux, S. Dynamic expression of the LIM-homeodomain gene Lhx15 through larval brain development of the sea lamprey (*Petromyzon marinus*). *Gene Expr. Patterns* **6**, 873–878 (2006).
68. Wang, J. *et al.* Tracing cell-type evolution by cross-species comparison of cell atlases. *Cell Rep.* **34**, 108803 (2021).
69. Suryanarayana, S. M. & Huilgol, D. Conservation and Diversification of Pallial Cell Types across Vertebrates: An Evo-Devo Perspective. *Brain. Behav. Evol.* **98**, 210–228 (2023).
70. Tosches, M. A. & Laurent, G. Evolution of neuronal identity in the cerebral cortex. *Curr. Opin. Neurobiol.* **56**, 199–208 (2019).
71. Arendt, D. *et al.* The origin and evolution of cell types. *Nat. Rev. Genet.* **17**, 744–757 (2016).
72. Doyle, J. J. Cell types as species: Exploring a metaphor. *Front. Plant Sci.* **13**, 1–23 (2022).
73. Achim, K. & Arendt, D. Structural evolution of cell types by step-wise assembly of cellular modules. *Curr. Opin. Genet. Dev.* **27**, 102–108 (2014).
74. Arendt, D. The evolution of cell types in animals: Emerging principles from molecular studies. *Nat. Rev. Genet.* **9**, 868–882 (2008).
75. Arendt, D. *et al.* The origin and evolution of cell types. *Nat. Rev. Genet.* **17**, 744–757 (2016).
76. Arendt, D., Bertucci, P. Y., Achim, K. & Musser, J. M. Evolution of neuronal types and families. *Curr. Opin. Neurobiol.* **56**, 144–152 (2019).
77. Tarashansky, A. J. *et al.* Mapping single-cell atlases throughout metazoa unravels cell type evolution. *Elife* **10**, 1–24 (2021).

78. Mah, J. L. & Dunn, C. W. Reconstructing cell type evolution across species through cell phylogenies of single-cell RNAseq data. *bioRxiv* (2023).
79. Geirsdottir, L. *et al.* Cross-Species Single-Cell Analysis Reveals Divergence of the Primate Microglia Program. *Cell* **179**, 1609-1622.e16 (2019).
80. Schlosser, G., Musser, J. & Arendt, D. Editorial - Development and evolution of sensory cells and organs. *Dev. Biol.* **431**, 1–2 (2017).
81. Arendt, D. Animal evolution: Of flame and collar cells. *Curr. Biol.* **31**, R1003–R1006 (2021).
82. Deryckere, A., Woych, J., Jaeger, E. C. B. & Tosches, M. A. Molecular Diversity of Neuron Types in the Salamander Amygdala and Implications for Amygdalar Evolution. *Brain. Behav. Evol.* **98**, 61–75 (2023).
83. Miyashita, T., Gess, R. W., Tietjen, K. & Coates, M. I. Non-ammocoete larvae of Palaeozoic stem lampreys. *Nature* **591**, 408–412 (2021).
84. Zeisel, A. *et al.* Molecular Architecture of the Mouse Nervous System. *Cell* **174**, 999-1014.e22 (2018).
85. Muhammad Tibi, A. *et al.* A telencephalon cell type atlas for goldfish reveals diversity in the evolution of spatial structure and cell types. *Sci. Adv.* **9**, 2023.06.19.545605 (2023).
86. Liguz-Lecznar, M. & Skangiel-Kramska, J. Vesicular glutamate transporters (VGLUTs): The three musketeers of glutamatergic system. *Acta Neurobiol. Exp. (Wars).* **67**, 207–218 (2007).
87. Lannoo, M. J. & Hawkes, R. A search for primitive Purkinje cells: zebrin II expression in sea lampreys (*Petromyzon marinus*). *Neurosci. Lett.* **237**, 53–55 (1997).
88. Kebschul, J. M. *et al.* Cerebellar nuclei evolved by repeatedly duplicating a conserved cell-type set. *Science (80-.).* **370**, (2020).
89. Smith, J. J. *et al.* Sequencing of the sea lamprey (*Petromyzon marinus*) genome provides insights into vertebrate evolution. *Nat. Genet.* **45**, 415–421 (2013).
90. Yu, D. *et al.* Hagfish genome illuminates vertebrate whole genome duplications and their evolutionary consequences. *bioRxiv* 2023.04.08.536076 (2023).
91. Weil, M. T. *et al.* Axonal ensheathment in the nervous system of lamprey: Implications for the evolution of myelinating glia. *J. Neurosci.* **38**, 6586–6596 (2018).
92. Bandler, R., Mayer, C. & Fishell, G. Cortical interneuron specification: the juncture of genes, time and geometry Rachel. *Curr. Opin. Neurobiol.* **42**, 17–24 (2017).
93. Chen, Y. J. J. *et al.* Single-cell RNA sequencing identifies distinct mouse medial

ganglionic eminence cell types. *Sci. Rep.* **7**, 1–11 (2017).

94. Cobos, I., Shimamura, K., Rubenstein, J. L. R., Martínez, S. & Puelles, L. Fate Map of the Avian Anterior Forebrain at the Four-Somite Stage, Based on the Analysis of Quail–Chick Chimeras. *Dev. Biol.* **239**, 46–67 (2001).
95. Cobos, I., Puelles, L. & Cobos, I. The avian telencephalic subpallium originates inhibitory neurons that invade tangentially the pallium (Dorsal Ventricular Ridge and Cortical Areas). *Dev. Biol.* **239**, 30–45 (2001).
96. Colquitt, B. M., Merullo, D. P., Konopka, G., Roberts, T. F. & Brainard, M. S. Cellular transcriptomics reveals evolutionary identities of songbird vocal circuits. *Science (80-.).* **371**, (2021).
97. Métin, C. *et al.* Conserved pattern of tangential neuronal migration during forebrain development. *Development* **134**, 2815–2827 (2007).
98. Jiménez, S. *et al. Analysis of pallial/cortical interneurons in key vertebrate models of Testudines, Anurans and Polypteriform fishes*. *Brain Structure and Function* **225**, (Springer Berlin Heidelberg, 2020).
99. Puelles, L., Alonso, A., García-Calero, E. & Martínez-de-la-Torre, M. Concentric ring topology of mammalian cortical sectors and relevance for patterning studies. *J. Comp. Neurol.* **527**, 1731–1752 (2019).
100. Bergsten, J. A review of long-branch attraction. *Cladistics* **21**, 163–193 (2005).
101. Simakov, O. *et al.* Deeply conserved synteny resolves early events in vertebrate evolution. *Nat. Ecol. Evol.* **4**, 820–830 (2020).
102. Zhao, C., Wang, C., Zhang, H. & Yan, W. A mini-review of the role of vesicular glutamate transporters in Parkinson's disease. *Front. Mol. Neurosci.* **16**, (2023).
103. Paganos, P., Voronov, D., Musser, J., Arendt, D. & Arnone, M. I. Single cell rna sequencing of the strongylocentrotus purpuratus larva reveals the blueprint of major cell types and nervous system of a nonchordate deuterostome. *Elife* **10**, 1–29 (2021).
104. Grillner, S. & Robertson, B. The Basal Ganglia Over 500 Million Years. *Curr. Biol.* **26**, R1088–R1100 (2016).
105. Ocaña, F. M. *et al.* The lamprey pallium provides a blueprint of the mammalian motor projections from cortex. *Curr. Biol.* **25**, 413–423 (2015).
106. Brox, A., Puelles, L., Ferreiro, B. & Medina, L. Expression of the genes Emx1, Tbr1, and Eomes (Tbr2) in the telencephalon of *Xenopus laevis* confirms the existence of a ventral pallial division in all tetrapods. *J. Comp. Neurol.* **474**, 562–577 (2004).
107. Garcia-Calero, E., Martínez-de-la-Torre, M. & Puelles, L. *A radial histogenetic model*

of the mouse pallial amygdala. *Brain Structure and Function* **225**, (Springer Berlin Heidelberg, 2020).

108. Northcutt, R. & Kicliter, E. Organization of the amphibian telencephalon. In Comparative Neurology of the Telencephalon. in (ed. Ebbesson, S. O. E.) 203–255 (Plenum New York, Publishing Corporation, 1980).
109. Zeng, H. What is a cell type and how to define it? *Cell* **185**, 2739–2755 (2022).
110. Clevers, H. *et al.* What is your conceptual definition of "cell type" in the context ofa mature organism? *Cell Syst.* **4**, 255–259 (2017).
111. Morris, S. A. The evolving concept of cell identity in the single cell era. *Development* **146**, dev169748 (2019).
112. Vergara, H. M. *et al.* Whole-body integration of gene expression and single-cell morphology. *Cell* **184**, 4819-4837.e22 (2021).
113. Wang, K. *et al.* African lungfish genome sheds light on the vertebrate water-to-land transition. *Cell* **184**, 1362-1376.e18 (2021).
114. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
115. Andrews, S. FastQC: a quality control tool for high throughput sequence data. (2019).
116. Smit, A., Hubley, R. & Green, P. RepeatMasker Open-4.0.
117. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
118. Hoff, K. J., Lomsadze, A., Borodovsky, M. & Stanke, M. Whole-Genome Annotation with BRAKER. *Methods Mol. Biol.* **1962**, 65–95 (2019).
119. Brůna, T., Hoff, K. J., Lomsadze, A., Stanke, M. & Borodovsky, M. BRAKER2: automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. *NAR Genomics Bioinforma.* **3**, lqaa108 (2021).
120. Hoff, K. J., Lange, S., Lomsadze, A., Borodovsky, M. & Stanke, M. BRAKER1: Unsupervised RNA-Seq-Based Genome Annotation with GeneMark-ET and AUGUSTUS. *Bioinformatics* **32**, 767–769 (2016).
121. Lomsadze, A., Ter-Hovhannisyan, V., Chernoff, Y. O. & Borodovsky, M. Gene identification in novel eukaryotic genomes by self-training algorithm. *Nucleic Acids Res.* **33**, 6494–6506 (2005).
122. Buchfink, B., Xie, C. & Huson, D. H. Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* **12**, 59–60 (2015).
123. Gotoh, O. A space-efficient and accurate method for mapping and aligning cDNA

sequences onto genomic sequence. *Nucleic Acids Res.* **36**, 2630–2638 (2008).

124. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
125. Barnett, D. W., Garrison, E. K., Quinlan, A. R., Strömberg, M. P. & Marth, G. T. BamTools: a C++ API and toolkit for analyzing and managing BAM files. *Bioinformatics* **27**, 1691–1692 (2011).
126. Lomsadze, A., Burns, P. D. & Borodovsky, M. Integration of mapped RNA-Seq reads into automatic training of eukaryotic gene finding algorithm. *Nucleic Acids Res.* **42**, e119 (2014).
127. Stanke, M., Diekhans, M., Baertsch, R. & Haussler, D. Using native and syntenically mapped cDNA alignments to improve de novo gene finding. *Bioinformatics* **24**, 637–644 (2008).
128. Stanke, M., Schöffmann, O., Morgenstern, B. & Waack, S. Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. *BMC Bioinformatics* **7**, 62 (2006).
129. Brůna, T., Lomsadze, A. & Borodovsky, M. GeneMark-EP+: eukaryotic gene prediction with self-training in the space of genes and proteins. *NAR genomics Bioinforma.* **2**, lqaa026 (2020).
130. Emms, D. M. & Kelly, S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* **20**, 238 (2019).
131. Marlétaz, F. *et al.* Amphioxus functional genomics and the origins of vertebrate gene regulation. *Nature* **564**, 64—70 (2018).
132. Satou, Y. *et al.* Improved genome assembly and evidence-based global gene model set for the chordate *Ciona intestinalis*: New insight into intron and operon populations. *Genome Biol.* **9**, 1–11 (2008).
133. Venkatesh, B. *et al.* Elephant shark genome provides unique insights into gnathostome evolution. *Nature* **505**, 174–179 (2014).
134. 25 genomes for 25 years (sanger.ac.uk/collaboration/25-genomes-for-25-years).
135. Braasch, I. *et al.* The spotted gar genome illuminates vertebrate evolution and facilitates human-teleost comparisons. *Nat. Genet.* **48**, 427–437 (2016).
136. Howe, K. *et al.* The zebrafish reference genome sequence and its relationship to the human genome. *Nature* **496**, 498–503 (2013).
137. Amemiya, C. T. *et al.* The African coelacanth genome provides insights into tetrapod evolution. *Nature* **496**, 311–316 (2013).

138. Hellsten, U. *et al.* The genome of the Western clawed frog *Xenopus tropicalis. Science* **328**, 633–636 (2010).
139. Hillier, L. ., Miller, W., Birney, E. & Al., E. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* **432**, 695–716 (2004).
140. The Genome Reference Consortium.
141. Katoh, K. & Standley, D. . MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772– 780 (2013).
142. Nguyen, L.-T., Schmidt, H. A., von Haeseler, A. & Minh, B. Q. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268–274 (2015).
143. Emms, D. M. & Kelly, S. STRIDE: Species Tree Root Inference from Gene Duplication Events. *Mol. Biol. Evol.* **34**, 3267–3278 (2017).
144. Sarropoulos, I. *et al.* Developmental and evolutionary dynamics of cis-regulatory elements in mouse cerebellar cells. *Science* **373**, (2021).
145. Lun, A. T. L. *et al.* EmptyDrops: Distinguishing cells from empty droplets in droplet-based single-cell RNA sequencing data. *Genome Biol.* **20**, 1–9 (2019).
146. Germain, P., Lun, A., Garcia Meixide, C., Macnair, W. & Robinson, M. Doublet identification in single-cell sequencing data using scDblFinder. *f1000research* (2022). doi:doi:10.12688/f1000research.73600.2.
147. Hao, Y. *et al.* Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573-3587.e29 (2021).
148. Hafemeister, C. & Satija, R. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biol.* **20**, 1–15 (2019).
149. Consortium, T. G. O. Gene Ontology: tool for the unification of biology. *Nat. Genet.* **25**, 25–29 (2000).
150. Welch, J. D. *et al.* Single-Cell Multi-omic Integration Compares and Contrasts Features of Brain Cell Identity. *Cell* **177**, 1873-1887.e17 (2019).
151. Korsunsky, I. *et al.* Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat. Methods* **16**, 1289–1296 (2019).
152. Suzuki, R. & Shimodaira, H. Pvclust: an R package for assessing the uncertainty in hierarchical clustering. *Bioinformatics* **22**, 1540–1542 (2006).
153. Lambert, S. A. *et al.* The Human Transcription Factors. *Cell* **172**, 650–665 (2018).

154. Ke, R. *et al.* In situ sequencing for RNA analysis in preserved tissue and cells. *Nat. Methods* **10**, 857–860 (2013).

155. Cheng, Y.-S. *et al.* Autonomous combinatorial color barcoding for multiplexing single molecule RNA visualization. *bioRxiv* 127373 (2017). doi:10.1101/127373