Essay



# Towards a Rational Terminology for Cell Types

Yuanzhen Zhu *, Guang Li, Peter W.H. Holland, Günter P. Wagner *, Sebastian M. Shimeld *

Posted Date: 12 June 2026

doi: 10.20944/preprints202606.1059.v1

Keywords: evolutionary cell biology; cell type nomenclature

 doi:10.20944/preprints202606.1059.v1



*Essay*

# Towards a Rational Terminology for Cell Types

**Yuanzhen Zhu [1,*], Guang Li [2], Peter W.H. Holland [1], Günter P. Wagner [3,4,*] and Sebastian M. Shimeld [1,*]**

[1] Department of Biology, University of Oxford, Oxford, OX1 3EL, UK.
[2] State Key Laboratory of Cellular Stress Biology, School of Life Sciences, Xiamen University, Xiamen, China.
[3] Department of Ecology and Evolutionary Biology, Yale University, USA.
[4] Department of Evolutionary Biology University of Vienna, Vienna, Austria.
* Correspondence: yuanzhen.zhu@biology.ox.ac.uk (Y.Z.); gunter.wagner@yale.edu (G.P.W.); sebastian.shimeld@biology.ox.ac.uk (S.M.S.)

**Abstract**

Cell types are fundamental biological units and partially independent evolutionary units, shaped by individualised gene regulatory networks and developmental lineages. Despite the recent explosion in single-cell sequencing and increased attention on cell characterization, we still lack a unified and consistent naming system for cell types that works across species. Since cell types are the products of evolutionary diversification, we propose that cell-type names should explicitly reflect evolutionary history, and suggest a naming system with a phylogenetic representation prefix as a simple, informative and intuitive way to do this. The key to this is establishing the evolutionary/taxonomic level of comparison, coupled with understanding homology and innovation in cell-type evolution. Put simply, it can apply to both individual cell types and their clades. We illustrate this approach using two case studies: chordate macroglia and more explicitly on vertebrate photoreceptors. The long-term goal is to stimulate progress towards a more coherent and informative language for cell-type identity and comparative analyses that is evolutionarily extendable as single-cell research proliferates across the tree of life.

**Keywords:** evolutionary cell biology; cell type nomenclature

## 1. Introduction

An ongoing quest for cell biology is to characterize and compare cell types within and across species. This has catalysed the establishment of the Brain Initiative Cell Census Network [1], the Human Cell Atlas [2], the international Biodiversity Cell Atlas (BCA) initiative [3], and programmes supported by the Chan Zuckerberg Initiative [4,5], among others. With rapid advances in single-cell technologies, we anticipate a surge of studies in the coming years aimed at defining and comparing cell types across diverse animal lineages. To achieve its goals, the field needs a framework for naming cells in a way that is simple to apply, consistent and informative, and adaptable such that new information can be easily integrated into an existing framework. Many fundamental concepts and questions can only be properly framed and addressed once we specify what entities are being compared and at which taxonomic or evolutionary scale.

Historically, cell types have been named based on a range of criteria including morphology, function, marker expression, physiology, developmental lineage, evolutionary origin and their discoverers. Some names are widely used while others are not: whether a name persists likely reflects the conjunction of its utility with historical contingency. Some names based on function and morphology appear simple and clear enough at first glance: for example neurons generate action potentials, receive input and produce output via synapses, and have processes [6]. However, biology is staggeringly diverse, and these features may not coexist in neurons in many cases. For instances, amacrine cell literally means "without an axon", unipolar neurons only have one process [7],

 doi:10.20944/preprints202606.1059.v1



amacrine cells, horizontal cells, and bipolar cells in the retina, and most neurons in *C. elegans* do not fire action potentials [8,9]. In addition, these features may also exist in other cell types: some epithelial cells in jellyfish generate action potentials [10], hair cells, Merkel cells, and astrocytes have synapse-like structures [11–13], and body wall muscle cells of *C. elegans* send long processes to motor neurons, rather than conversely [14].

In most cases these situations reflect the evolutionary lability of features that might otherwise appear definitive. However, they raise the issue that a naming system does not accommodate diversity may compromise understanding of relationships, and potentially knowledge transfer. To date, consensus on a system that can do this for cell types is lacking [15]. This situation parallels the early history of gene naming, when different research groups assigned distinct names to what were later recognised as the 'same gene' or gene family in different species. For example, the gene *Small eye* in mouse, *Aniridia* in human, and *eyeless* and *twin of eyeless* in fly were identified over 30 years ago and through analyses such as molecular phylogenetics have been collectively united as co-orthologous members of a *Pax6* gene family [16,17]. Systematic analysis of genes from different species in an explicit evolutionary framework is now routine in genomics (for example through ENSEMBL gene trees and OrthoFinder) and vital for integrating information between species and studies.

While the situation with cells may parallel that of gene relationships at one level, it is much more complex and nuanced. Genes can be reduced to linear character sequences that can be relatively easily identified, compared and modelled with explicit evolutionary theory. In contrast, cell types are better compared to species than genes, in that both represent complex, integrated biological organisations rather than discrete units. Cell types are high-dimensional and quasi-independent [18] (partially independent [19]) characters, embedded within regulatory systems that interact continuously with other cells and the environment. They cannot be defined by a single marker or sequence, but instead by an integrated regulatory architecture—including core Transcription Factors (TFs), additional regulatory molecules, protein–protein interactions, chromatin states, signalling dependencies, effector genes, morphology, and developmental context. Unlike genes, which are directly encoded in the genome, cell types must be re-established through developmental processes in each generation. This added layer of complexity makes cross-species comparisons inherently more challenging. We therefore need multiple levels of evidence to support homology relationships for characters.

In recent years, a number of researchers have sought to understand the relationships between cell types within and across species [19–29]. Most have applied pairwise comparison rather than explicit phylogenetic structuring. Some practical cell-type 'phylogeny' methods have been proposed to date [30–32]. However, none are yet grounded in theories of cell-type evolution, which emphasise that core regulatory programmes more accurately capture evolutionary relationships among cell types. The field is currently transitioning from analyses largely confined to single species toward large-scale, comparative studies of cell-type evolution, mirroring the development from the early genomic era to comparative genomics. In this context, the need for a standardised, evolutionarily informed nomenclature for cell types is urgent.

In this perspective, we review the evolutionary context and nature of cell types. We pay particular attention to the hierarchical nature of the cell-type concept, as we consider this fundamental to defining relationships of cells between species. We then propose a nomenclature system capturing the level of phylogenetic representation for cell-type identity that address key issues and helps refine definitions. Finally, we apply this nomenclature to test cases to explore its benefits and limitations.

## 2. Evolutionary Theories and the Hierarchical Nature of Cell Types

Cell type, as a partially independent evolutionary unit, must use a specific combination of genomic and regulatory information which is not used by other cell types [19]. Here, partial independence refers to the fact that cell types remain genomically and developmentally integrated yet can evolve through partially individuated regulatory programmes. An evolutionary definition of

 doi:10.20944/preprints202606.1059.v1



cell types can be underpinned by common descent regardless of form and function [19,33]. In other words, cell types with very similar functions may not be homologous. For example, the Calvin cycle is performed in mesophyll cells in C3 plants and by bundle sheath cells in C4 plants, which are considered non-homologous cell types despite having the same critical function [34]. However, homology is a layered/hierarchical concept and can be different at different levels: the classic textbook anatomical example is bird and bat wings, which are homologous as forelimbs "but not as wings" (though note this is oversimplified and not entirely accurate, see Box1). In a similar sense cell type is also a hierarchical concept. For example, myocyte is a cell type, but so too are smooth and striated myocytes (including cardiac and skeletal myocytes), which are all nested within muscle/myocyte [35]. To explore this, we need to consider how cell types are specified and evolve.

**Box1: 1. character identity vs. character state.**

Statements about bat and bird wing homology are common as textbook examples but usually simplistic to the point of not being correct. The homology concept, introduced by Richard Owen nearly 200 years ago, is clear. Here we quote: "the same organ in different animals under every variety of form and function" [33]. The organ here is a *character identity*, it can be an organ, a tissue, or different (individualised) cell types. The "same" means historical continuity of information encoding the development of the character identity. "Under every variety of form and function" means *character states* can be quite different regarding function and morphology. The distinction between character identity and character state has been conceptually and genetically formalised in Wagner's Character Identity Networks framework [36,37]. This framework is supported by extensive empirical evidence across diverse species. Within this, bat wings and bird wings should not be treated as distinct characters, but rather as different character states of the vertebrate forelimbs. The developmental and genetic information underlying the forelimb predates both bats and birds and is shared across jawed vertebrates. Therefore, it is conceptually misleading to describe bat wings and bird wings as homologous to each other; instead, they represent divergent states of the same homologous forelimb character. For more detail about character identity and character state see references [36–38].

Inspired by the concept of biological homology [39], several studies have investigated and conceptualised the molecular and mechanistic basis for cell-type identity as governed by relatively conserved and self-sustained core regulatory networks. Concepts include Character Identity Networks (ChINs) [36,37], terminal selectors [40], and Core Regulatory Complexes (CoRCs) of TF proteins [19]. It is worth noting that CoRCs need not only include TFs. Other aspects such as coregulators [41], cofactors [42–44], non-coding RNAs [45], chromatin remodelers [46], post-translational modifications [47,48], protein-protein interactions [40,49], and several other regulatory mechanisms [48,50] that maintain the cell-type identity are also important. These concepts can help us to trace the "sameness" (historical continuity) of cell-type identity. More precisely, it is the historical continuity of key gene regulatory networks (conserved TFs which encode identity) rather than the expression of individual homologous genes that underlies the homology of characters [36].

According to the sister cell-type model, cell types can evolve by duplication and divergence [19,21], recombination/co-option [51] (cell-type 'fusion' [19]), and parallel recruitment (serial homology: the serial sister cell-type model [19]). More specifically, the sister cell-type model assumes that novel cell types evolve by duplication and divergence of cell-type-specific CoRCs or some other identity mechanisms [52] (Figure 1). The serial sister cell-type model suggests that new cell types can evolve by recruitment and individuation into a different developmental lineage or body parts (Figure 1). Recombination/co-option describes repurposing of existing regulatory programmes into existing cell-type-specific modules from which they were formerly absent (Figure 1). As recently suggested, co-option better refers to the reuse of developmental mechanisms or GRNs in non-homologous traits, not in the duplication of serial homologues [53]. Here, therefore, we use parallel recruitment rather than co-option to describe the process of evolving serially homologous cell types. Unless there is rampant recombination/co-option during cell-type evolution, novel cell types primarily arise through 'duplication and divergence' or parallel recruitment and are, in theory, expected to exhibit a tree

 doi:10.20944/preprints202606.1059.v1



structure. This is exactly what we see in studies of normal cell types [30,54,55]. Cases of co-option of CoRCs would impose a limited amount of network-structure or fusing of branches within a tree, analogous to hybridisation and introgression in phylogenetic trees of species. The tree structure of cell types highlights the hierarchical nature of cell-type homology. A given cell type may have different sister cell types depending on the level of phylogenetic representation considered. As an analogy, we can consider sister cell types as paralogues within a gene family that has duplicated repeatedly at different times in evolution, generating multiple genes. There would be multiple levels of paralogue relationships, each defined by the genes being compared and the level of the duplication(s) that separated them. Thus, the concept of sister cell types—defined as *two cell types that arise from the splitting and individuation of an ancestral cell type* [19]—should be specified with reference to the phylogenetic level at which duplication and divergence occurred. For example, astrocytes and ependymal cells can be viewed as sister cell types at the level of Vertebrata, as each was individuated before the radiation of this taxonomic group. Similarly astrocytes and oligodendrocytes are sister cell types at Gnathostomata level. Astrocytes in vertebrate brains also show heterogeneity and regional variance like neurons [56]. If novel astrocyte subtypes arose by duplication and divergence in primate ancestors, they would be sister cell types at the level of Primates. Without specifying the phylogenetic level all these cell types could be considered sister cell types as they do ultimately trace back to a common origin, making the concept vague.

It is particularly important to distinguish the concepts of innovation and evolution as defined in this context. Evolution here means descent with the same character identity, while innovation is the origin of novel characters. A novel character is a one that is neither homologous to any character in the ancestral species nor serially homologous to any other character of the same organism [37]. In brief, innovation = non-homology [37] (or specifically non-special homology and non-serial homology). Here we refine the concept of innovation as non-special homology or individuated serial homology. Therefore novel cell types can be defined with phylogenetic representation, *as long as they are individuated and traceable, and are always apomorphic or synapomorphic characters*. Individuation of serial homologues is probably the easiest and most frequent pathway to innovation of characters (organs, tissues, and cell types, etc). Without considering innovation = non-special homology or individuated serial homology, and commitment to what entities and CoRCs we are comparing, the homology concept becomes useless. For example, if we assume a single origin of neurons, then any neuronal subtypes in one species are homologous to any other neuronal type in another species. This would be true at a very high level, but it is not useful for many scientific questions.

Some studies have also noted a relationship between innovation and co-option that provides regulatory complexity by redeploying GRNs [53,57]. Because this can repurpose complex ‘preassembled’ genetic circuitry into new networks and contexts, it is thought to facilitate the emergence of novel characters. However, to be informative, it is necessary to clarify in detail how much co-option contributes to novelty, and at which biological level or unit this contribution occurs [53]. Co-option can occur at the level of effector genes or key TFs of CoRCs, but only the latter can potentiate alternative developmental outcomes that may lead to the emergence of a novel character (cell type) [58,59]. For example, appendages in vertebrates and arthropods originated independently and both co-opted an outgrowth programme likely present in the urbilaterian [60–62]; scarab beetle horns co-opted appendage outgrowth programmes [53,63,64]; a vertebrate first gill programme was co-opted into outer-ear development [65]; neural crest evolution coupled the co-option of several key TFs to the neural plate border of the vertebrate ancestor [66–69]. In contrast, duplication-and-divergence and parallel recruitment involve stepwise tweaks, and can relatively easily be traced back to the same evolutionary lineages [19].

 doi:10.20944/preprints202606.1059.v1



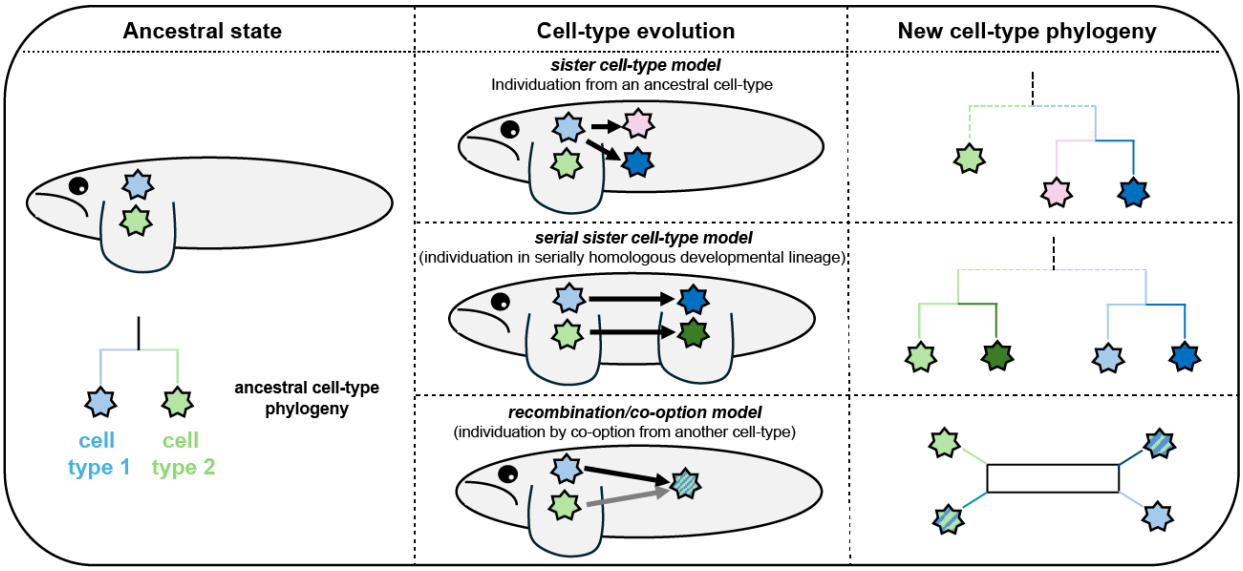


**Figure 1. | Cell-type evolutionary models.** The left column shows the ancestral states of an individualised developmental lineage, the middle column illustrates models of cell-type evolution, and the right column presents the cell-type phylogenies inferred under each model. An individuated developmental lineage does not represent a single lineage but instead comprises multiple indistinguishable, repeated lineages. In the serial sister cell-type model, when an ancestral lineage develops to two cell types (for example, 1 and 2) duplicated via parallel recruitment, evolutionary relationships among these cell types do not mirror their developmental relatedness [19].

## 3. A Cell-Type Nomenclature Reflecting the Evolutionary Taxonomy of Cell-Type Identity

The hierarchical nature and conceptual differences between evolution (the same character identity) and innovation (origin of novel characters) highlight the significance of comparing cell types at a suitable level within and across species. Researchers often use a variety of terms, such as family, subfamily, type, subtype, and subtype with specific markers, to describe different levels of cell-type hierarchy [70,71]. Such an approach may be informative within a single tissue/organ of a given species. However, these labels lack consistent, operational definitions across studies, organs, and species. As a result, no universal criteria exist to determine when two cell populations should be considered as distinct “families” or “types”, and this may eventually lead to the question of distinguishing between cell types and cell states (variants from the same cell type) [72]. This ambiguity limits the comparability of cell-type annotation across datasets and undermines cross-species analyses. It further highlights the need for an evolutionary framework for defining cell identity while explicitly capturing phylogenetic levels, as this accommodates core concepts of innovation, homology, hierarchy, and comparability. The need for an evolutionary taxonomy of cell types is also rooted in the fact that cell types are the product of evolution, as discussed above, and taxonomy should accommodate the causal processes that led to cell-type diversity.

Taking an example from above, if we assume all neurons share a common origin, then neurons in one species are homologous to those in another at the level of Eumetazoa. However, at finer resolutions many neuronal subtypes are not homologous between all species but instead are clade- or species-specific innovations (Figure 2a). For example, cnidocytes and vertebrate Purkinje cells are both neurons in a broad sense, yet they are non-homologous neuronal subtypes (Figure 2a). Assigning a Cnidaria prefix to cnidocytes immediately encodes the fact: it denotes a neuronal cell type that originated on the stem lineage of Cnidaria, specifies that clade, and is non-homologous to any other neuronal types in other animal groups. While such distinctions may appear obvious in this case, they become substantially less straightforward when cell types diverge more subtly with

 doi:10.20944/preprints202606.1059.v1



decreasing evolutionary scales. We therefore suggest four steps to perform comparative analysis and name cell types:

1. Define the comparative resolution and taxonomic depth at which homology is biologically meaningful (and relevant to your scientific questions), starting broadly and avoiding over-resolution.
2. Test this 'resolution' using comparative cell atlases and methods to assess whether cross-species correspondences can be established at this level.
3. Infer homology by identifying conserved core regulatory programmes—such as conserved combinations of key TFs—evaluated within a phylogenetic context rather than by pairwise similarity alone.
4. Assign cell-type names by placing phylogenetic prefixes at the deepest clade supported by the conserved regulatory identity, and define innovations as lineage-restricted divergences from ancestral programmes (or which do not exist in the outgroup).

We exemplify the above steps using a case from our recent study [29]. Step1: we wanted to study common cell-type repertoires in the vertebrate brain. Step 2: we found clear and strong cross-species mapping for astrocytes and ependymal cells across vertebrates, and oligodendrocytes across gnathostomes but not cyclostomes (Figure 2b). Step 3: we then identified key TFs conserved across vertebrates for each glial type. By examining the expression patterns of these key TFs in amphioxus—which do not show specific co-expression as in vertebrates—we were able to confirm their phylogenetic origins within the Vertebrata. Step 4: we named the cell types according to their phylogenetic distribution as Vertebrata astrocytes, Vertebrata ependymal cells and Gnathostomata oligodendrocytes.

 doi:10.20944/preprints202606.1059.v1



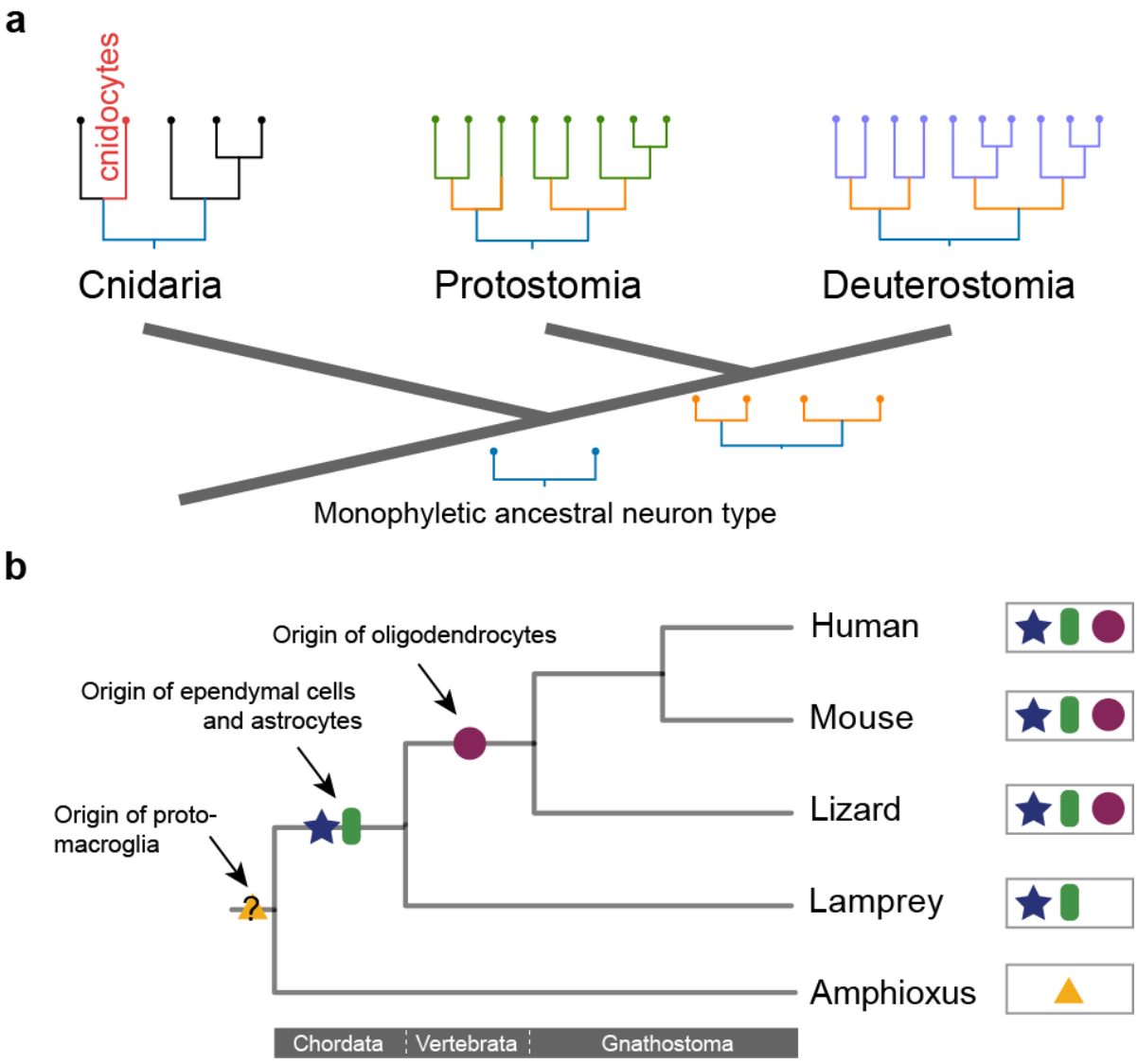


**Figure 2. | Inferring macroglia innovation and homology across species and assigning taxonomic prefix.** a, Hypothetical scheme illustrating neuron diversification in animal evolution. This illustration is adapted from Ref. [73]. In each tree, cell-type branching events that had already occurred in a given ancestor are indicated by the colour assigned to that ancestor. Older cell-type families are shared across distantly related phyla, whereas younger families are restricted to individual phyla or even to narrower taxonomic levels. b, inferred phylogenetic position for the origin of macroglia subtypes (astrocytes, ependymal cells, oligodendrocytes).

As we previously highlighted, functional definition of cell types can be meaningful in closely related species but is problematic in wider phylogenetic comparisons. As we explore below, human "green cones" and "blue cones" are *functionally similar* to their counterparts in chickens but are not homologous. Instead, human "green cones" evolved from ancestral red cones and are thus homologous to chicken red cones [74] and "blue cones" spectrally shifted from ancestral ultraviolet cones and thus homologous to chicken ultraviolet cones. Additionally at a deeper evolutionary level, all photoreceptors share a common origin and are thus homologous at a higher phylogenetic level [75]. Therefore, a clear description and definition of what we are comparing, within an explicit phylogenetic context, is extremely important to identify clear innovations and trace historical continuity (homology) during evolution.

To describe cell types more explicitly and clearly across species, we are not trying to abandon current cell-type names. Many are deeply embedded in the literature and in pervasive day-to-day usage. Instead, we propose introducing a phylogenetic prefix that reflects evolutionary origins.

 doi:10.20944/preprints202606.1059.v1



Specifically, adding prefix for an individualised cell type that corresponds to the highest taxonomic level that can be traced.

## 4. Examples of Cell-Type Cases with the New Nomenclature

Several studies have addressed the origin and evolution of some of the major cell types. Cnidaria cnidocytes [76] are discussed above. Below we consider some of these and show how phylogenetic naming would work in these cases. We use italics to indicate new proposed names.

1). *Vertebrata astrocytes* and *Vertebrata ependymal cells* are multifunctional macroglial cells essential to CNS function. They are specified by conserved core regulators (*Sox9, Rfx4,* and *Olig2* for astrocytes, and *Sox9*, *Foxj1*, and *Msx1* for ependymal cells) and display comparable developmental and molecular features across vertebrates [24,29,77]. *Gnathostomata oligodendrocytes* are myelinating glial cells critical for rapid nerve conduction in the CNS. They arise from neural progenitors and are specified by conserved TFs – including *Sox8/10*, *Prox1*, *Olig1/2*, *Nkx2.2*, *Nkx6.2*, *Ceb5*, *Atf2* – shared across several jawed vertebrates [24,29,78]. While amphioxus possesses glial cells highly expressing *SoxE*, *Eaat2*, *Notch*, *Ife1*, *Ifn2*, and *Ife2a*, no glial cells in adult amphioxus CNS specifically co-express the above key TFs (instead many key TFs are broadly expressed). Similarly, while tunicates possess glial cells expressing neurotransmitter transporters (*Eaats*) and *Glul* (glutamine synthetase) [79,80], no glial cells in tunicate or amphioxus larvae specifically co-express these core TFs. These data suggest an ancestral Chordata glial cell type, and hence that astrocytes and ependymal cells are vertebrate-level innovations. Although several key myelin structural genes (*Mbp*, *Mog*, *Mag*, etc) in jawed vertebrates are absent in jawless vertebrates [81], this alone is insufficient to support oligodendrocyte cell identity as originating in the ancestor of gnathostomes [78]. We recently analysed the expression of oligodendrocyte conserved TFs and conserved TFs separating astrocytes and oligodendrocytes [29]. None of these core regulators were found to be specifically co-expressed in glial populations in jawless vertebrates at either larval or adult stages, supporting the view that oligodendrocytes represent a gnathostome innovation.

2*). Pterygota Kenyon cells* (KCs) are intrinsic neurons of the mushroom bodies, playing a central role in learning, memory, and sensory integration. Although a detailed phylogenetic comparison relaying on single-cell brain datasets of insects is still lacking, the origin of KCs likely coincides with that of the mushroom bodies, which can be traced at least back to Pterygota. KCs arise from neuroblasts during embryonic and larval development and are specified by key TFs (e.g., *eyeless, trio*) [23,82]. Some basal insect lineages (e.g., mayflies) possess mushroom bodies, whereas others among the most "basal" groups (such as the wingless Archaeognatha) lack them [83]. While mushroom body-like structures have been identified in some crustaceans and even velvet worms [84,85], the origin of KCs across insects or perhaps panarthropods remain unsolved.

3). *Eutheria Invasive trophoblast* and *Eutheria Predecidual cells* are highly specialised cell types that define placental mammals and underpin extraplacental invasive placentation and sustained pregnancy. Trophoblasts form the foetal component of the placenta and mediate implantation and foetal–maternal exchange, with invasive behaviour evolving in eutherians (though subsequently lost in several lineages [86]). Recently, opossum trophoblast giant cells have been proposed as homologous to the invasive trophoblasts of eutherians, suggesting that a trophoblast cell-type family (*Theria trophoblast*) originated in the common ancestor of therians [28]. Analyses indicate shared invasive gene and conserved TF (ADAMTS18, GATA3, PAPPA2, QSOX1, SCAI, and ZEB1) [28,87,88] programmes of *Eutherian invasive trophoblast*. The *Theria Decidual stromal cell* family underwent stepwise evolution into other two types [28]. Specifically, *Theria Decidual stromal cells* are conserved across therian mammals and evolved into *Eutheria predecidual cells*, characterised by the expression of MEIS1, WT1, ESR1, PGR, EGFR, and HAND2, and subsequently to *Euarchontoglires endocrine decidual stromal cells*, marked by PRL, DCN, PPIB, and DUSP1 [28].

4). *Theria barrier-associated* (*border-associated*) *macrophages* (BAMs) are responsible for immune surveillance and scavenge waste alongside microglia in mammalian brain [89]. They are functionally similar to zebrafish vascular-derived mural lymphatic endothelial cells (muLEC), but are non-

 doi:10.20944/preprints202606.1059.v1



homologous based on differences in key TFs (OSR2, LOXL1, PROX1 vs. SPI1 and IRF8) and ontogeny [90].

5). *Vertebrata neural crest cells* are transient, multipotent, and migratory, enabling the development of vertebrate innovations like craniofacial bones, gill arches, facial structures, and sensory ganglia. As such they are distinct form the cells discussed above in that they are a transient developmental cell type destined to form an array of terminally differentiated cell types. They arise from the neural plate border and are specified by key regulatory genes (*SoxE*, *Foxd3*, *Snai1/2*, *Tfap2*, and *Twist*) that are not expressed in the lateral edges of the neural plate of amphioxus and tunicates [67]. Hence, the taxonomic prefix "Vertebrata" for this cell-type family.

## 5. A Detailed Example: Photoreceptor Evolution and Nomenclature in Vertebrates

The above examples illustrate how a phylogenetically informed nomenclature could apply to some familiar cell types that have been relatively well studied. To explore further how this could work in detail, we now turn to photoreceptors as these have been the subject of intensive work by many researchers, meaning we have deeper understanding of photoreceptor cells and gene expression across different phylogenetic scales. The molecular machinery underlying photoreceptors and phototransduction is deeply conserved, indicating that most animal photoreceptors—and their non-visual derivatives—share a common bilaterian origin [20,21,91,92]. Photoreceptors are broadly divided into ciliary and rhabdomeric types [93], which differ in morphology, opsins, signalling pathways, and regulatory programmes, and are present in both protostomes and deuterostomes [94,95]. Insects primarily use rhabdomeric photoreceptors for vision, organised into ommatidia, whereas vertebrates rely on ciliary photoreceptors (rods and cones) that signal through complex retinal circuits. Rhabdomeric-like cells are also present in vertebrates but largely serve non-visual roles: for example vertebrate retinal interneurons may have evolved from ancestral rhabdomeric photoreceptors [20,21,96]. Additionally, the evolutionary relatedness among photoreceptor subtypes remain incompletely understood [97]. Current nomenclature systems reflect functional partitioning: insect and vertebrate photoreceptors and subtypes are largely named based on morphology, opsin spectral sensitivity, or the opsins they express [94,95,98].

Understanding of vertebrate photoreceptors has been significantly enhanced by recent comparative, molecular, and developmental studies and we therefore focus on these. A recent photoreceptor nomenclature framework [99] provides an excellent summary of key regulators and established consensus subtype names that correct prior misuse. However, like many existing frameworks, it does not explicitly define the hierarchical relationships among cell types, which limits its applicability across broader evolutionary contexts and to other cell types. Here, we outline a framework for tracing and naming vertebrate photoreceptors, using the hierarchical taxonomic approach.

### *5.1. Vertebrata Rods and Vertebrata Cones*

As outlined in our four-steps above, we begin with broader classifications. Rods and cones are two main types of ciliary photoreceptors that have been identified in the retina of jawed vertebrates and sea lamprey and are specified by a set of conserved TFs [75]. In particular, *Nrl* and *Nr2e3* defines rods, while *Thrb2* and *Rxrg* marks cones in vertebrates [75]. Many genes in the phototransduction cascade between rods and cones are ohnologues, for example *Gnat1/2*, *Gnb1/3*, *Gngt1/2*, *Pde6a/b/c*, *Gnga1/3*, *Gbgb1/3*, *Grk1/7*, and *Sag*/*Arr3* [100], a conclusion supported by paralogon and phylogeny analyses [100]. Therefore rods and cones are sister cell types at the level of Vertebrata. Two recent studies [29,101] surveyed amphioxus adult brain but only identified rhabdomeric photoreceptors, which highly express *Otp*, *Islet*, *Lim1/5*, and *melanopsin* (r-opsin gene) (data not shown). Given that many key TFs and effectors underlying vertebrate rods and cones arose through duplication following whole genome duplication (WGD), and that many of the key TFs involved in rods and

 doi:10.20944/preprints202606.1059.v1



cones are not involved in *Ciona* photoreceptor specification (see markers table in [102]), it is most likely that amphioxus and tunicates do not possess vertebrate-style rods and cones. Then we can consider rods and cones are vertebrate innovations and can be renamed as *Vertebrata Rods* and *Vertebrata Cones*.

It is worth noting that all photoreceptors (and many other cells) from the retina develop from retinal progenitor cells, so the close development and evolutionary relationships of vertebrate photoreceptors aligns with the sister cell type model (at least for the early stage of duplication and divergence). Cones have several subtypes (generally classified by opsin expression, see details below), however, a single cone cluster expresses only *LWS* (long-wave-sensitive opsin gene) (other cone opsins are either lost or pseudogenes) in the sea lamprey [75]. Short-wave-sensitive opsin gene (*SWS* family: *Rh1*/*Rh2*/*SWS2*/*SWS1*) and *LWS* genes were long thought to arise through tandem duplication prior to WGDs [100]. Following updates to our understanding of the timing of WGDs in vertebrates [103,104] the evolutionary history of opsin genes has been revised, with additional genomes and analyses particularly in lamprey. This identified *LWS* and *Rh1*, *SWS2* and *Rh2* in two paralogons duplicated by the first vertebrate WGD (1R) and conserved across lampreys [105], suggesting the vertebrate ancestor already had tandemly duplicated *LWS*, *SWS1*, *SWS2*, and *Rh* prior to 1R.

Additionally, photoreceptors in lamprey are more morphologically cone-like and five types were identified in the pouched lamprey *Geotria australis* from morphology viewpoint [106]. We might particularly need single-cell data from several lamprey species (especially the pouched lamprey) to fully resolve cone diversity and origin. Specifically, if Red-, UV-, Blue-, and Green-cones can be found in lamprey, these lineages should all be assigned a Vertebrata prefix. Interestingly, the loss of photoreceptors and opsins in vertebrates is non-random [98]: Rods, Red-cones, *Rh1*, and *LWS* are generally retained, whereas Blue- and Green-cones, as well as *SWS2*, and *Rh2* have been independently lost in multiple lineages (Figure 3). Together with the central role of Rods and Red-cones in supporting vision [98], this pattern offers an additional perspective for a vertebrate ancestral photoreceptor repertoire likely comprising Rods and Red-cones, consistent with the observations in the sea lamprey [75]. For now, we conservatively apply this Vertebrata prefix only to the overall cone group reflecting this uncertainty.

### *5.2. Gnathostomata Red-, Blue-, Green-, and UV-Cones, and Amniota/Tetrapoda Principle and Accessory Double Cones*

We now focus on cones. Although single-cell retina datasets remain limited, current evidence shows opsin and cell-type identity are often coupled [98,107] (but not always, e.g., mouse and other rodent Red-cones co-express *LWS* & *SWS1* [108,109], some fish hypothesised Red-cones instead use *Rh2* [98] while a few fish have lost *Rh2* and express a green-shifted *LWS* instead [110], and in squamates morphologically rod-like photoreceptors expressing cone opsins and vice versa [111]). To describe these cone subtypes, ancestral states and descendants are here capitalised (e.g., Green-cones), whereas photoreceptor types defined by spectral properties are written in lowercase and enclosed in quotation marks (e.g., “green cones”). Double-cones consist of two tightly associated cones – a principal double cone (DC-P) and an accessory double cone (DC-A).

Two cone subtypes, including Red-cones and UV-cones, have been found in human, macaque, mouse, and squirrels, corresponding to *LWS*/*MWS* and *SWS1* opsin expression, respectively [74]. Recent retinal datasets from chicken and brown anole contain all six cone subtypes [74,112] while zebrafish contain all four major single-cones [74]. In addition double cones are identified morphologically in amphibians and platypus [113–115]. Although single-cell atlases of frogs and platypus are still lacking, together, these observations support two inferences. First, DC-P and DC-A were likely innovations that arose in the tetrapod lineage (or at least in the stem lineage of amniotes) [74,112]. Second, the major single-cone classes (Red-, Blue-, Green-, and UV-cones) as well as Rods can be at least traced back to gnathostome ancestors [74,112] (Figure 3).

 doi:10.20944/preprints202606.1059.v1



We performed a similar approach to that used in our recent study [29] to find conserved key transcriptional regulators (TFs, coregulators, and cofactors) for each photoreceptor and to understand the sister cell type of DC-P and DC-A, using the datasets from Tommasini and colleagues [74]. We systemically identified previous reported key transcriptional regulators, for example, NRL/MAF, NR2E3, ESRRB, SAMD11 for Rods [116–120], THRB, SOX5 for Red-cones [121], TBX2 for UV-cones [122], NFIA/B/C/X, FOXQ2 for Blue-cones [74,123], NFIA/B/C/X for Green-cones [74], THRB, SAMD7 for DC-P, and SKOR1 for DC-A. Meanwhile, we also found other conserved key regulators including RORB, CBFA2T2, CITED2, and YBX3 for Rods, ISL2 for Red-cones, EGR1, RORA, LHX4, SKOR1 for UV-cones, TBX2, LHX4, SKOR1 for Blue-cones, and CRIP2 for DC-A. More importantly, we found many key conserved TFs are not very 'binarized', meaning they are often highly and specifically expressed by more than one photoreceptor subtype. This may reflect shared regulatory mechanisms in photoreceptor differentiation, such as temporal control of thyroid hormone (TH) signalling. For example, dynamic expression of TH-degrading and -activating proteins regulates TH levels, with low signalling specifying UV-cones and high signalling promoting "L/M cones" [124] ("L/M cones" are Red-cones, see below for details).

We can therefore name the six cone subtypes as *Gnathostomata Red-*, *Blue-*, *Green-*, *UV-cones*, and *Tetrapoda/Amniota DC-P* and *DC-A.* Additional platypus, birds, reptiles, frogs, and salamanders retinal datasets are required to better understand the conserved programmes separating derived DC-P and DC-A from ancestral Red-cones and Blue-cones [112], respectively (or possibly both from Red-cones: see [74]). Therefore Red-, Blue-, Green, UV-cones are sister cell types at least at the level of Gnathostomata, DC-P and Red-cones are sister cell types at Tetrapoda/Amniota, DC-A and Red-cones/Blue-cones are sister cell types at Tetrapoda/Amniota.

Now we take a closer look at the situation of cell-type loss. Character loss is frequent in evolution, and any cell-type naming strategy needs to accommodate this. Common ancestors of eutherians were likely dichromats, having lost two opsins during early mammalian evolution: *Rh2* (ancestral green-sensitive) lost in Therians and *SWS2* (ancestral blue-sensitive) lost in Eutherians [125] (Figure 3). Trichromatic vision re-evolved independently in Old World monkeys and in some New World monkeys, through different genetic mechanisms: gene duplication of the *LWS* (red-sensitive) into *MWS* (green-sensitive) in Old World monkeys, and X-chromosomal opsin polymorphisms in some New World species [125,126]. Three isolated photoreceptors, including Red-cones, UV-cones, and Rods have been found in human, macaque, mouse, and squirrels, corresponding to *LWS/MWS*, *SWS1*, and *Rh1* opsin expression, respectively [74]. Data from opossums (a marsupial) show that they clearly possess all the above three cell types (with Double-cones) [74,126,127].

Platypus (monotremes) have undergone lineage-specific losses of *Rh2* and *SWS1* while retaining *Rh1*, *SWS2*, and *LWS* [125]. Although photoreceptor single-cell datasets from platypus are lacking, the platypus genome retains *Foxq2*, encoding a TF conserved in Blue-cone specification that has been lost in placental mammals [112]. We therefore infer the ancestral photoreceptor repertoires of mammals likely include Red-cones, Green-cones, Blue-cones, UV-cones, Rods, and Double-cones, with the loss of Blue-cones in therians and Green-cones and Double-cones in eutherians (Figure 3). Consistent with this and as mentioned above, data from chicken and lizards reveal distinct Red-cones, Green-cones, Blue-cones, UV-cones, Rods, and Double-cones [74] (Figure 3).

Therefore for example, previously named "M- and S-cones" (based on wavelength-sensitivity) in mice [128] should be named as Red- and UV-cones from an evolutionary perspective [129]. Similarly, *LWS* gene is named as *MWS* in mice based on spectrum, which could be misleading for readers and researchers outside of evolutionary biology. These also highlight the evolvability of spectrum and lineage-specific gain and loss of opsin and colour visions in mammals, suggesting that using spectral sensitivity is not ideal to understand either opsin relationships or cell-type identity of photoreceptors.

 doi:10.20944/preprints202606.1059.v1



*5.3. Are There Catarrhini (Old World Primates) Novel Cone Subtypes?*

Interestingly, "red cones" (or "L cones") and "green cones" (or "M cones") in macaque map to a single cluster and are only distinguishable by the expression of *LWS* or *MWS* (which was duplicated from *LWS* in the ancestors of Old World Primates) [129]. Two opsins are almost identical with only ~ 2% differences on coding sequences and no other differentially expressed genes were identified between "red and green cones" in macaque. This may reflect Red-cones with the same cell-type identity making a stochastic choice between neighbouring *LWS* and *MWS* opsin genes [129]. However, we still lack extensive human and other primate retinal datasets capturing enough cone cells and diversity [130]. Consequently, the potential genetic (e.g., CoRC) and developmental differences (if there are any) between the "red and green cones" in Old World Primates remain unresolved, as does the question of whether "green cones" represent a novel cell-type associated with an individualised CoRC; at present, this appears unlikely. We therefore need larger and more comprehensive datasets to confirm species/clade-specific cell types. However, distinguishing cell types from cell states across species at high resolution requires special and careful considerations (see discussion).

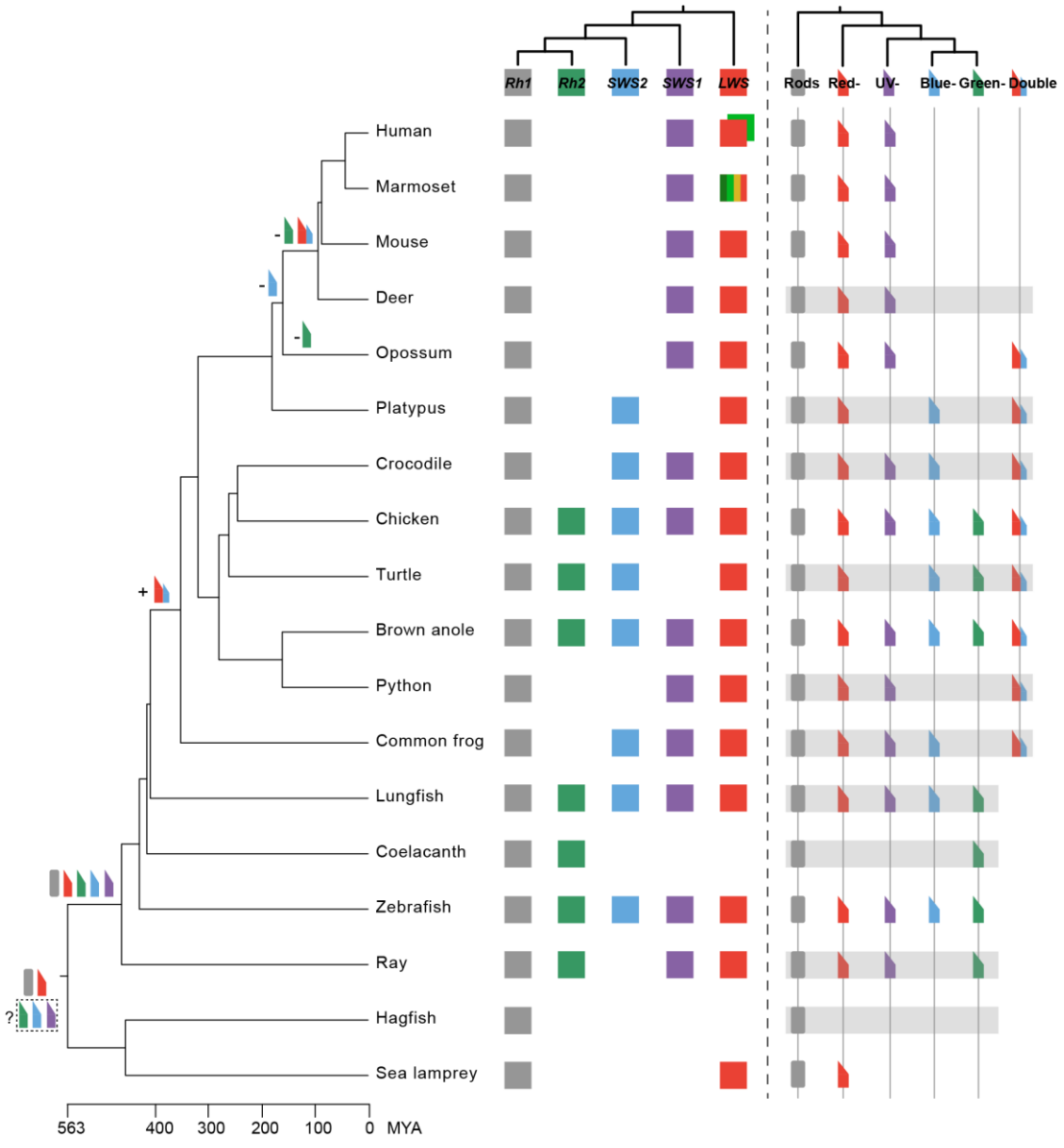


**Figure 3. | Vertebrate opsin and photoreceptor repertoires.** Phylogeny of some representative vertebrates and maximal known opsin and photoreceptor cell-type repertoires. The opsin cladogram is based on Collin et al [131]

 doi:10.20944/preprints202606.1059.v1



(please refer to Ref. [95,105] for discussion and recent findings on opsin phylogeny). The opsin gene repertoires are based on compiled data from Hagen and colleagues [132] and gene trees we built on orthogroups. The photoreceptor cell identities are based on several photoreceptor studies cited in the main text. The semi-transparent grey boxes indicate missing or insufficient data, and therefore cell-type assignments are inferred from opsin expression patterns. Inferred gain and loss of photoreceptors are shown on the species phylogeny. Please note that the photoreceptor phylogeny here is inferred based on key conserved regulators and might be revised in future studies with more species and datasets. Furthermore, the species included in this analysis may not fully represent the diversity of the entire clade. We show one possibility that DC-P and DC-A are derived from Red and Blue-cones respective, but an alternative possibility is both of them are derived from Red-cones. MYA, million years ago.

## 6. Cell-Type Radiation Occurred at Vertebrate Stem Lineage?

We have discussed several clear cell-type evolutionary lineages, but using species evolution as an analogy, one may ask whether there are episodes of rapid cell-type radiation analogous to adaptive radiation, in which a single ancestral species diversified into many descendant species within a relatively short evolutionary timescale. The striking differences between invertebrate chordates and vertebrates marked by the origin of an extensive number of novel cell types may be the case. Another group and us found that amphioxus/*Ciona* cell types lacking clear one-to-one homologs in vertebrates (something that has been referred to as 'diffuse homology') [29,133]. Dopaminergic neurons, long thought to be homologous across chordates [26,134], are in fact homologous between amphioxus and *Ciona* but not to vertebrate dopaminergic neurons [29]. Based on these observations, three hypotheses can be proposed to explain this intriguing phenomenon:

1). The cell type without clear homology across phylum/subphylum may have originated from different ancestral cell types.

2). The CoRCs or terminal selectors could be less conserved than previously observed and expected.

3). Some cell-type families, such as neural families, greatly expanded and diversified independently in different phyla/subphyla, blurring evolutionary histories.

In *C. elegans*, combinations of homeodomain TFs are sufficient to specify all neuronal classes [135]. In chordates and vertebrates, however, the situation is likely more complex. In our opinion, the third hypothesis is most likely the case for the diffuse homology in brain cell types between amphioxus and vertebrates, as we can clearly identify broader glia family shared across chordates in our study (Figure 4) [29]. Still, the first two may provide to a certain extent explanation in other cases. It might be necessary to integrate specific (potentially irreversible) genomic and regulatory apomorphy together with evidence about TF interactions and post translational modifications, i.e., CoRCs, to better understand and identify cell types [136].

 doi:10.20944/preprints202606.1059.v1



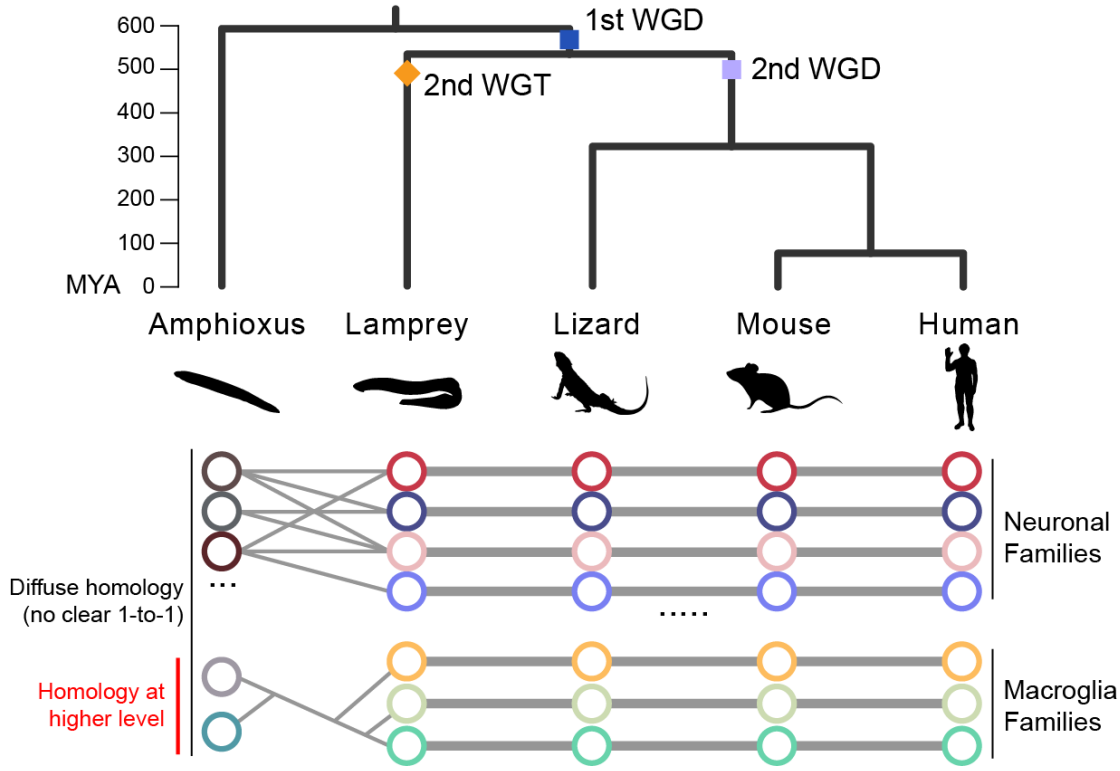


**Figure 4. | Illustration of cell-type relationships between amphioxus and vertebrates.** Summary of key conclusions of Ref. [29]. MYA, million years ago.

## 7. Strengths, Limitations, and Future Directions

The nomenclature we propose not only provides a consistent naming system for biologists across diverse fields, it also establishes an explicitly evolutionary and comparative framework for understanding cell-type origins, diversification, and homology. By encoding phylogenetic 'time of origin' information directly into cell-type names it simplifies future comparative analyses: researchers can immediately infer the evolutionary level at which a given cell type originated as an innovation. This, in turn, reframes comparison questions from whether cell types are related to how their evolutionary histories unfold—for example, whether the inferred origin can be traced further back in the tree, or whether subsequent lineage-specific losses have occurred. This compels researchers to be explicit as to what is being compared, what are the key conserved regulators, and at what phylogenetic level. Put more bluntly, it forces us to be more careful in how we invoke and study homology and innovation. Important concepts (e.g., sister cell types, innovation, different levels of cell types) can be more precisely defined with a phylogenetic level.

A challenge is understanding individuation and discrete/continuous identity and state. Transcriptomic differences may reflect either complete individuation, incomplete individuation, or simply condition-dependent variation between cells (the cell state). The keys to distinguish complete and incomplete individuation could be irreversibility and evolutionary constraints on CoRCs [19], and potential genomic architectures and regulatory entanglement [136–138]. The specification of cells is controlled by a combination of intrinsic and extrinsic factors. How these contribute to cell type and cell state has yet to be fully elucidated in different cell types. We might particularly need comprehensive cell-type profiles from multiple time points and under different conditions (perhaps even from experimentally transplanted cells). For example, broad and distinct 'families' of cortical inhibitory neurons (e.g., Vip, Pvalb, Sst, Lamp5, Sncg) exhibit largely non-overlapping phenotypes and clearly distinct transcriptomic profiles [139]. In contrast, clusters within the same 'family' show a continuum of variability in both phenotype and expression level [139]. The extent to which this variability is shaped by intrinsic and extrinsic factors, and whether it reflects continuous subtype identities, remains unclear. We also need to understand variation between individuals within a

 doi:10.20944/preprints202606.1059.v1



species population to understand the influence of environmental factors and genetic differences on cell types and cell states [72].

Another challenge is resolving cell types at increasingly fine resolution. Recent work has shown that incorporating regional identity improves the efficiency of trans-differentiation between astrocytes and neurons [56]. With higher resolution, cell types are more likely to be identifiable as species- or clade-specific, and these subtypes may play distinct roles in species/clade-specific physiology and disease. Conceptually, this raises questions about how character individuality is realised. Even duplicated genes with identical sequences are treated as distinct entities because they occupy different genomic loci. In contrast, repeated and nearly identical cells—such as erythrocytes or hair matrix keratinocytes—are typically classified as a single cell type because they are not individualised [19,37]. This distinction underscores the need for clearer criteria when understanding and defining cell types at higher resolution.

Cell types must be re-established in every generation, therefore understanding the evolution of developmental lineages is also key to explaining the emergence of novel cell types. It is worth noting that developmental lineage does not necessarily recapitulate evolutionary process and is actually often incongruent [19]. Still, more developmental and comparative studies are required. Extending analyses from cell type to tissue level is a promising direction, as biological organisation and evolution operate across multiple scales: molecules, cell types, circuits, territories, tissues, organs. These levels can evolve partially independently, and homology at one level does not imply conservation at others [140].

A long-term goal is to compare and name cell types automatically in a manner similar or close to genes. A central and challenging step is the reconstruction of cell-type phylogeny. Achieving this will require substantially more multimodal datasets, comparisons across diverse species, close species, individuals within population, different tissues and organs, and the development of new/extended theoretical and computational frameworks to better understand and ultimately model the evolutionary and mechanistic processes that generate and diversify cell types.

**Terminology section**

1. Special homology (or homology): the same character in different animals under every variety of form and function.
2. Serial homology: repetitive (but individuated) character in the same organism.
3. Cell-type homology: cell types that trace back to the same cell type defined with phylogenetic representation.
4. Core regulatory complex (CoRC): A protein complex composed of terminal selector transcription factors that enables and maintains the identity of cell types.
5. Novel cell types: individuated and evolutionarily traceable cell types defined with phylogenetic representation. Phylogenetic representation denotes the origin of individuated cell types and related CoRCs.
6. Apomorphic character: a specialized or derived character unique to a specific species or group, distinguishing them from their ancestors.
7. Synapomorphic character: an apomorphic character shared by two or more taxa and is therefore hypothesised to have evolved in their most recent common ancestor.
8. Paralogon: a set of paralogous chromosomal regions derived from a common ancestral region through ancient gene or whole-genome duplication events.
9. Diffuse homology: no clear one-to-one homology, likely derived from independent radiation at certain cell-type families.
10. Regulatory entanglement: Persistent co-localization of genes and regulatory elements emerged through genomic rearrangements and the integration of multiple enhancer–promoter interactions, thereby constraining subsequent evolution.

**Acknowledgments:** We thank Diego Dolgetta-Garcia, Thomas D. Lewin, Sisi Chen, Mei Wang, and Alessandro Pennati for helpful discussions and suggestions. We are grateful to Dario Tommasini and Karthik Shekhar for

 doi:10.20944/preprints202606.1059.v1



providing the pre-processed single-cell atlases used in their recent publication. We thank Desheng Duan for constructive feedback on the figure aesthetic design.

**Author Contributions:** Y.Z. conceived the study, wrote the original draft, and conducted the analysis. Y.Z., G.P.W., and S.M.S. conceptualised the theories. Y.Z., G.L., P.W.H.H., G.P.W., and S.M.S contributed to evaluation and investigation and revised the manuscript.

**Competing interests:** The authors declare no competing interests.

**Data availability:** The preprocess atlases are deposited in the figshare repository under the link: https://doi.org/10.6084/m9.figshare.31850833.

**Code availability:** All scripts associated with this analysis are publicly available in GitHub repository under the link: https://github.com/DiracZhu1998/Perspective_analysis_for_Vertebrate_Photoreceptors.

## References:

1. Bakken, T. E. et al. Comparative cellular analysis of motor cortex in human, marmoset and mouse. *Nature* **598**, 111–119 (2021).
2. Regev, A. et al. The Human Cell Atlas. *Elife* **6**, e27041 (2017).
3. Sebé-Pedrós, A. et al. The Biodiversity Cell Atlas: mapping the tree of life at cellular resolution. *Nature* **645**, 877–885 (2025).
4. Pearce, J. D. et al. A Cross-Species Generative Cell Atlas Across 1.5 Billion Years of Evolution: The TranscriptFormer Single-cell Model. Preprint at https://doi.org/10.1101/2025.04.25.650731 (2025).
5. CZI Cell Science Program et al. CZ CELLxGENE Discover: a single-cell data platform for scalable exploration, analysis and modeling of aggregated data. *Nucleic Acids Research* **53**, D886–D900 (2025).
6. Kristan, W. B. Early evolution of neurons. *Current Biology* **26**, R949–R954 (2016).
7. *The Oxford Handbook of Invertebrate Neurobiology*. (Oxford University Press, New York, NY, 2019).
8. Neves, G. & Lagnado, L. The retina. *Current Biology* **9**, R674–R677 (1999).
9. Goodman, M. B., Hall, D. H., Avery, L. & Lockery, S. R. Active Currents Regulate Sensitivity and Dynamic Range in C. elegans Neurons. *Neuron* **20**, 763–772 (1998).
10. Meech, R. W. Electrogenesis in the lower Metazoa and implications for neuronal integration. *J Exp Biol* **218**, 537–550 (2015).
11. Glowatzki, E., Grant, L. & Fuchs, P. Hair cell afferent synapses. *Curr Opin Neurobiol* **18**, 389–395 (2008).
12. Hoffman, B. U. et al. Merkel Cells Activate Sensory Neural Pathways through Adrenergic Synapses. *Neuron* **100**, 1401-1413.e6 (2018).
13. Allen, N. J. & Eroglu, C. Cell Biology of Astrocyte-Synapse Interactions. *Neuron* **96**, 697–708 (2017).
14. Altun, Zeynep F., Hall, David H., & Herndon, Laura A. Muscle System - Introduction. https://doi.org/10.3908/WORMATLAS.1.6 doi:10.3908/WORMATLAS.1.6.
15. Domcke, S. & Shendure, J. A reference cell tree will serve science better than a reference cell atlas. *Cell* **186**, 1103–1114 (2023).
16. Feiner, N., Meyer, A. & Kuraku, S. Evolution of the Vertebrate Pax4/6 Class of Genes with Focus on Its Novel Member, the Pax10 Gene. *Genome Biology and Evolution* **6**, 1635–1651 (2014).
17. Quiring, R., Walldorf, U., Kloter, U. & Gehring, W. J. Homology of the *eyeless* Gene of *Drosophila* to the *Small eye* Gene in Mice and *Aniridia* in Humans. *Science* **265**, 785–789 (1994).
18. Lewontin, R. C. Adaptation. *Sci Am* **239**, 212–230 (1978).
19. Arendt, D. et al. The origin and evolution of cell types. *Nat Rev Genet* **17**, 744–757 (2016).
20. Arendt, D. Evolution of eyes and photoreceptor cell types. *Int J Dev Biol* **47**, 563–571 (2003).
21. Arendt, D. The evolution of cell types in animals: emerging principles from molecular studies. *Nat Rev Genet* **9**, 868–882 (2008).
22. Tarashansky, A. J. et al. Mapping single-cell atlases throughout Metazoa unravels cell type evolution. *eLife* **10**, e66747 (2021).
23. Li, Q. et al. A single-cell transcriptomic atlas tracking the neural basis of division of labour in an ant superorganism. *Nat Ecol Evol* **6**, 1191–1204 (2022).

 doi:10.20944/preprints202606.1059.v1




24. Lamanna, F. et al. A lamprey neural cell type atlas illuminates the origins of the vertebrate brain. *Nat Ecol Evol* https://doi.org/10.1038/s41559-023-02170-1 (2023) doi:10.1038/s41559-023-02170-1.
25. Church, S. H., Mah, J. L. & Dunn, C. W. Unification of species, gene, and cell trees for single-cell expression analyses. Preprint at https://doi.org/10.48550/ARXIV.2307.02561 (2023).
26. Dai, Y. et al. Evolutionary origin of the chordate nervous system revealed by amphioxus developmental trajectories. *Nat Ecol Evol* https://doi.org/10.1038/s41559-024-02469-7 (2024) doi:10.1038/s41559-024-02469-7.
27. Li, Y. et al. Origin and stepwise evolution of vertebrate lungs. *Nat Ecol Evol* https://doi.org/10.1038/s41559-025-02642-6 (2025) doi:10.1038/s41559-025-02642-6.
28. Stadtmauer, D. J. et al. Cell type and cell signalling innovations underlying mammalian pregnancy. *Nat Ecol Evol* https://doi.org/10.1038/s41559-025-02748-x (2025) doi:10.1038/s41559-025-02748-x.
29. Zhu, Y. et al. Whole genome duplication shaped cell-type evolution in the vertebrate brain. *Nature* https://doi.org/https://doi.org/10.1038/s41586-026-10629-x (2026) doi:https://doi.org/10.1038/s41586-026-10629-x.
30. Mah, J. L. & Dunn, C. W. Cell type evolution reconstruction across species through cell phylogenies of single-cell RNA sequencing data. *Nat Ecol Evol* https://doi.org/10.1038/s41559-023-02281-9 (2024) doi:10.1038/s41559-023-02281-9.
31. Price, P. D. et al. Detecting signatures of selection on gene expression. *Nat Ecol Evol* **6**, 1035–1045 (2022).
32. Tosches, M. A. et al. Evolution of pallium, hippocampus, and cortical cell types revealed by single-cell transcriptomics in reptiles. *Science* **360**, 881–888 (2018).
33. Owen, R. & Royal College of Physicians of Edinburgh. *Lectures on the Comparative Anatomy and Physiology of the Invertebrate Animals : Delivered at the Royal College of Surgeons, in 1843* /. (Longman, Brown, Green, and Longmans, London :, 1843). doi:10.5962/bhl.title.113849.
34. Swift, J. et al. Exaptation of ancestral cell-identity networks enables C4 photosynthesis. *Nature* **636**, 143–150 (2024).
35. Brunet, T. et al. The evolutionary origin of bilaterian smooth and striated myocytes. *eLife* **5**, e19607 (2016).
36. Wagner, G. P. The developmental genetics of homology. *Nat Rev Genet* **8**, 473–479 (2007).
37. Wagner, G. P. *Homology, Genes, and Evolutionary Innovation*. (Princeton University Press, 2014). doi:10.23943/princeton/9780691156460.001.0001.
38. DiFrisco, J., Love, A. C. & Wagner, G. P. The hierarchical basis of serial homology and evolutionary novelty. *Journal of Morphology* **284**, (2023).
39. Wagner, G. P. The Biological Homology Concept. *Annu. Rev. Ecol. Syst.* **20**, 51–69 (1989).
40. Hobert, O. Terminal Selectors of Neuronal Identity. in *Current Topics in Developmental Biology* vol. 116 455–475 (Elsevier, 2016).
41. Volkov, L. I. et al. *Samd7* represses short-wavelength cone genes to preserve long-wavelength cone and rod photoreceptor identity. *Proc. Natl. Acad. Sci. U.S.A.* **121**, e2402121121 (2024).
42. Haberle, V. et al. Transcriptional cofactors display specificity for distinct types of core promoters. *Nature* **570**, 122–126 (2019).
43. Joshi, K., Lee, S., Lee, B., Lee, J. W. & Lee, S.-K. LMO4 Controls the Balance between Excitatory and Inhibitory Spinal V2 Interneurons. *Neuron* **61**, 839–851 (2009).
44. Loker, R., Sanner, J. E. & Mann, R. S. Cell-type-specific Hox regulatory strategies orchestrate tissue identity. *Current Biology* **31**, 4246-4255.e4 (2021).
45. Long, Y., Wang, X., Youmans, D. T. & Cech, T. R. How do lncRNAs regulate transcription? *Sci. Adv.* **3**, eaao2110 (2017).
46. Osuma, A. et al. Cell type-specific functions of the PBAF chromatin-remodeling complex in neuronal diversification. *Nat Commun* **17**, 741 (2025).
47. Filtz, T. M., Vogel, W. K. & Leid, M. Regulation of transcription factor activity by interconnected post-translational modifications. *Trends in Pharmacological Sciences* **35**, 76–85 (2014).
48. Hobert, O. Gene Regulation by Transcription Factors and MicroRNAs. *Science* **319**, 1785–1786 (2008).
49. Nnamani, M. C. et al. A Derived Allosteric Switch Underlies the Evolution of Conditional Cooperativity between HOXA11 and FOXO1. *Cell Reports* **15**, 2097–2108 (2016).
50. He, M. et al. Cell-Type-Based Analysis of MicroRNA Profiles in the Mouse Brain. *Neuron* **73**, 35–48 (2012).

 doi:10.20944/preprints202606.1059.v1




51. Schlosser, G. *Evolutionary Origin of Sensory and Neurosecretory Cell Types*. (CRC Pess, Taylor & Francis Group, Boca Raton London New York, 2021).
52. DiFrisco, J., Love, A. C. & Wagner, G. P. Character identity mechanisms: a conceptual model for comparative-mechanistic biology. *Biol Philos* **35**, 44 (2020).
53. DiFrisco, J., Wagner, G. P. & Love, A. C. Reframing research on evolutionary novelty and co-option: Character identity mechanisms versus deep homology. *Seminars in Cell & Developmental Biology* **145**, 3–12 (2023).
54. Kin, K., Nnamani, M. C., Lynch, V. J., Michaelides, E. & Wagner, G. P. Cell-type Phylogenetics and the Origin of Endometrial Stromal Cells. *Cell Reports* **10**, 1398–1409 (2015).
55. Liang, C., the FANTOM Consortium, Forrest, A. R. R. & Wagner, G. P. The statistical geometry of transcriptome divergence in cell-type evolution and cancer. *Nat Commun* **6**, 6066 (2015).
56. Herrero-Navarro, Á. et al. Astrocytes and neurons share region-specific transcriptional signatures that confer regional identity to neuronal reprogramming. *Sci. Adv.* **7**, eabe8978 (2021).
57. Schlosser, G. Rebuilding ships while at sea—Character individuality, homology, and evolutionary innovation. *Journal of Morphology* **284**, (2023).
58. Erwin, D. H. & Davidson, E. H. The evolution of hierarchical gene regulatory networks. *Nat Rev Genet* **10**, 141–148 (2009).
59. Pomaville, M. B., Sattler, S. M. & Abitua, P. B. A new dawn for the study of cell type evolution. *Development* **151**, (2024).
60. Shubin, N., Tabin, C. & Carroll, S. Deep homology and the origins of evolutionary novelty. *Nature* **457**, 818–823 (2009).
61. Panganiban, G. et al. The origin and evolution of animal appendages. *Proc. Natl. Acad. Sci. U.S.A.* **94**, 5162–5166 (1997).
62. Shubin, N., Tabin, C. & Carroll, S. Fossils, genes and the evolution of animal limbs. *Nature* **388**, 639–648 (1997).
63. Hu, Y., Linz, D. M. & Moczek, A. P. Beetle horns evolved from wing serial homologs. *Science* **366**, 1004–1007 (2019).
64. Moczek, A. P. Integrating micro- and macroevolution of development through the study of horned beetles. *Heredity* **97**, 168–178 (2006).
65. Thiruppathy, M. et al. Repurposing of a gill gene regulatory program for outer-ear evolution. *Nature* **639**, 682–690 (2025).
66. Martik, M. L. et al. Evolution of the new head by gradual acquisition of neural crest regulatory circuits. *Nature* **574**, 675–678 (2019).
67. Martik, M. L. & Bronner, M. E. Riding the crest to get a head: neural crest evolution in vertebrates. *Nat Rev Neurosci* **22**, 616–626 (2021).
68. Meulemans, D. & Bronner-Fraser, M. Gene-Regulatory Interactions in Neural Crest Evolution and Development. *Developmental Cell* **7**, 291–299 (2004).
69. Meulemans, D. & Bronner-Fraser, M. Central role of gene cooption in neural crest evolution. *J. Exp. Zool.* **304B**, 298–303 (2005).
70. Tadross, J. A. et al. A comprehensive spatio-cellular map of the human hypothalamus. *Nature* **639**, 708–716 (2025).
71. Steuernagel, L. et al. HypoMap—a unified single-cell gene expression atlas of the murine hypothalamus. *Nat Metab* **4**, 1402–1419 (2022).
72. Zeng, H. What is a cell type and how to define it? *Cell* **185**, 2739–2755 (2022).
73. Arendt, D., Bertucci, P. Y., Achim, K. & Musser, J. M. Evolution of neuronal types and families. *Current Opinion in Neurobiology* **56**, 144–152 (2019).
74. Tommasini, D., Yoshimatsu, T., Puthussery, T., Baden, T. & Shekhar, K. Comparative transcriptomic insights into the evolution of vertebrate photoreceptor types. *Current Biology* **35**, 2228-2239.e4 (2025).
75. Wang, J. et al. Molecular characterization of the sea lamprey retina illuminates the evolutionary origin of retinal cell types. *Nat Commun* **15**, 10761 (2024).

 doi:10.20944/preprints202606.1059.v1




76. Babonis, L. S., Enjolras, C., Ryan, J. F. & Martindale, M. Q. A novel regulatory gene promotes novel cell fate by suppressing ancestral fate in the sea anemone *Nematostella vectensis*. *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2113701119 (2022).
77. Rowitch, D. H. & Kriegstein, A. R. Developmental genetics of vertebrate glial–cell specification. *Nature* **468**, 214–222 (2010).
78. Hines, J. H. Evolutionary Origins of the Oligodendrocyte Cell Type and Adaptive Myelination. *Front. Neurosci.* **15**, 757360 (2021).
79. Esposito, R. Comprehensive characterization of glial cells in the urochordate Ciona intestinalis. (2021).
80. Medina, B. N. S. P. et al. Identification of astrocyte-like cells in an adult ascidian during regeneration of the central nervous system. *Glia* **72**, 2190–2200 (2024).
81. Nave, K.-A. & Werner, H. B. Ensheathment and Myelination of Axons: Evolution of Glial Functions. *Annu. Rev. Neurosci.* **44**, 197–219 (2021).
82. Davie, K. et al. A Single-Cell Transcriptome Atlas of the Aging Drosophila Brain. *Cell* **174**, 982-998.e20 (2018).
83. Strausfeld, N. J., Sinakevitch, I., Brown, S. M. & Farris, S. M. Ground plan of the insect mushroom body: Functional and evolutionary implications. *J of Comparative Neurology* **513**, 265–291 (2009).
84. Strausfeld, N. J., Wolff, G. H. & Sayre, M. E. Mushroom body evolution demonstrates homology and divergence across Pancrustacea. *eLife* **9**, e52411 (2020).
85. Martin, C. et al. The velvet worm brain unveils homologies and evolutionary novelties across panarthropods. *BMC Biol* **20**, 26 (2022).
86. Mika, K., Whittington, C. M., McAllan, B. M. & Lynch, V. J. Gene expression phylogenies and ancestral transcriptome reconstruction resolves major transitions in the origins of pregnancy. *eLife* **11**, e74297 (2022).
87. Yang, X. & Meng, T. MicroRNA-431 affects trophoblast migration and invasion by targeting ZEB1 in preeclampsia. *Gene* **683**, 225–232 (2019).
88. Brandt, D. T. et al. SCAI acts as a suppressor of cancer cell invasion through the transcriptional control of β1-integrin. *Nat Cell Biol* **11**, 557–568 (2009).
89. Herz, J., Filiano, A. J., Wiltbank, A. T., Yogev, N. & Kipnis, J. Myeloid Cells in the Central Nervous System. *Immunity* **46**, 943–956 (2017).
90. Gaudi, A. U. et al. Convergent evolution of scavenger cell development at brain borders. *Nature* https://doi.org/10.1038/s41586-025-10003-3 (2026) doi:10.1038/s41586-025-10003-3.
91. Arendt, D., Tessmar-Raible, K., Snyman, H., Dorresteijn, A. W. & Wittbrodt, J. Ciliary Photoreceptors with a Vertebrate-Type Opsin in an Invertebrate Brain. *Science* **306**, 869–871 (2004).
92. Arendt, D., Tessmar, K., De Campos-Baptista, M.-I. M., Dorresteijn, A. & Wittbrodt, J. Development of pigment-cup eyes in the polychaete *Platynereis dumerilii* and evolutionary conservation of larval eyes in Bilateria. *Development* **129**, 1143–1154 (2002).
93. Eakin, R. M. Evolution of Photoreceptors. *Cold Spring Harbor Symposia on Quantitative Biology* **30**, 363–370 (1965).
94. Roberts, N. S., Hagen, J. F. D. & Johnston, R. J. The diversity of invertebrate visual opsins spanning Protostomia, Deuterostomia, and Cnidaria. *Developmental Biology* **492**, 187–199 (2022).
95. Lamb, T. D. Evolution of phototransduction, vertebrate photoreceptors and retina. *Progress in Retinal and Eye Research* **36**, 52–119 (2013).
96. Kafetzis, G., Bok, M., Baden, T. & Nilsson, D. A median eye origin of the vertebrate retina explains its unique circuitry. Preprint at https://doi.org/10.1101/2025.09.11.675609 (2025).
97. Hahn, J. et al. Evolution of neuronal cell classes and types in the vertebrate retina. *Nature* **624**, 415–424 (2023).
98. Baden, T. Ancestral photoreceptor diversity as the basis of visual behaviour. *Nat Ecol Evol* **8**, 374–386 (2024).
99. Baden, T. et al. A standardized nomenclature for the rods and cones of the vertebrate retina. *PLoS Biol* **23**, e3003157 (2025).
100. Larhammar, D., Nordström, K. & Larsson, T. A. Evolution of vertebrate rod and cone phototransduction genes. *Phil. Trans. R. Soc. B* **364**, 2867–2880 (2009).
101. Lin, C.-Y. et al. A 3D Amphioxus Brain Atlas Illuminates the Blueprint of the Ancestral Chordate Brain. Preprint at https://doi.org/10.1101/2025.11.27.690639 (2025).

 doi:10.20944/preprints202606.1059.v1




102. Cao, C. et al. Comprehensive single-cell transcriptome lineages of a proto-vertebrate. *Nature* **571**, 349–354 (2019).
103. Marlétaz, F. et al. The hagfish genome and the evolution of vertebrates. *Nature* **627**, 811–820 (2024).
104. Yu, D. et al. Hagfish genome elucidates vertebrate whole-genome duplication events and their evolutionary consequences. *Nat Ecol Evol* https://doi.org/10.1038/s41559-023-02299-z (2024) doi:10.1038/s41559-023-02299-z.
105. Lagman, D., Bergqvist, C. A. & Kuraku, S. Tandem Gene Clusters as Phylogenetic Anchors Reveal the Hidden History of Vertebrate Visual Opsins. *Molecular Biology and Evolution* **42**, msaf231 (2025).
106. Collin, S. P., Davies, W. L., Hart, N. S. & Hunt, D. M. The evolution of early vertebrate photoreceptors. *Phil. Trans. R. Soc. B* **364**, 2925–2940 (2009).
107. Baden, T. From water to land: Evolution of photoreceptor circuits for vision in air. *PLoS Biol* **22**, e3002422 (2024).
108. Lukáts, A. et al. Visual pigment coexpression in all cones of two rodents, the Siberian hamster, and the pouched mouse. *Invest Ophthalmol Vis Sci* **43**, 2468–2473 (2002).
109. Applebury, M. L. et al. The Murine Cone Photoreceptor. *Neuron* **27**, 513–523 (2000).
110. Musilova, Z. & Cortesi, F. The evolution of the green-light-sensitive visual opsin genes (RH2) in teleost fishes. *Vision Research* **206**, 108204 (2023).
111. Simões, B. F. et al. Multiple rod–cone and cone–rod photoreceptor transmutations in snakes: evidence from visual opsin gene expression. *Proc. R. Soc. B.* **283**, 20152624 (2016).
112. Liu, Y. et al. Avian photoreceptor homologies and the origin of double cones. *Current Biology* **35**, 2215-2227.e6 (2025).
113. Zeiss, C. J., Schwab, I. R., Murphy, C. J. & Dubielzig, R. W. Comparative retinal morphology of the platypus. *Journal of Morphology* **272**, 949–957 (2011).
114. Gao, F., Pang, J. & Wu, S. M. Sign-preserving and sign-inverting synaptic interactions between rod and cone photoreceptors in the dark-adapted retina. *The Journal of Physiology* **591**, 5711–5726 (2013).
115. Schott, R. K. et al. Diversity and Evolution of Frog Visual Opsins: Spectral Tuning and Adaptation to Distinct Light Environments. *Mol Biol Evol* **41**, msae049 (2024).
116. Kubo, S. et al. Functional analysis of Samd11, a retinal photoreceptor PRC1 component, in establishing rod photoreceptor identity. *Sci Rep* **11**, 4180 (2021).
117. Onishi, A. et al. The orphan nuclear hormone receptor *ERR* β controls rod photoreceptor survival. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 11579–11584 (2010).
118. Mears, A. J. et al. Nrl is required for rod photoreceptor development. *Nat Genet* **29**, 447–452 (2001).
119. Chen, J., Rattner, A. & Nathans, J. The Rod Photoreceptor-Specific Nuclear Receptor Nr2e3 Represses Transcription of Multiple Cone-Specific Genes. *J. Neurosci.* **25**, 118–129 (2005).
120. Oh, E. C. T. et al. Rod differentiation factor NRL activates the expression of nuclear receptor NR2E3 to suppress the development of cone photoreceptors. *Brain Research* **1236**, 16–29 (2008).
121. Ng, L. et al. A thyroid hormone receptor that is required for the development of green cone photoreceptors. *Nat Genet* **27**, 94–98 (2001).
122. Alvarez-Delfin, K. et al. Tbx2b is required for ultraviolet photoreceptor cell specification during zebrafish retinal development. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 2023–2028 (2009).
123. Ogawa, Y., Shiraki, T., Fukada, Y. & Kojima, D. Foxq2 determines blue cone identity in zebrafish. *Sci Adv* **7**, eabi9784 (2021).
124. Eldred, K. C. et al. Thyroid hormone signaling specifies cone subtypes in human retinal organoids. *Science* **362**, eaau6348 (2018).
125. Hagen, J. F. D., Roberts, N. S. & Johnston, R. J. The evolutionary history and spectral tuning of vertebrate visual opsins. *Developmental Biology* **493**, 40–66 (2023).
126. Jacobs, G. H. Evolution of colour vision in mammals. *Phil. Trans. R. Soc. B* **364**, 2957–2967 (2009).
127. Ebeling, W., Natoli, R. C. & Hemmi, J. M. Diversity of Color Vision: Not All Australian Marsupials Are Trichromatic. *PLoS ONE* **5**, e14231 (2010).
128. Euler, T., Haverkamp, S., Schubert, T. & Baden, T. Retinal bipolar cells: elementary building blocks of vision. *Nat Rev Neurosci* **15**, 507–519 (2014).

 doi:10.20944/preprints202606.1059.v1




129. Peng, Y.-R. et al. Molecular Classification and Comparative Taxonomics of Foveal and Peripheral Cells in Primate Retina. *Cell* **176**, 1222-1237.e22 (2019).
130. Menon, M. et al. Single-cell transcriptomic atlas of the human retina identifies cell types associated with age-related macular degeneration. *Nat Commun* **10**, 4902 (2019).
131. Collin, S. P., Davies, W. L., Hart, N. S. & Hunt, D. M. The evolution of early vertebrate photoreceptors. *Phil. Trans. R. Soc. B* **364**, 2925–2940 (2009).
132. Hagen, J. F. D., Roberts, N. S. & Johnston, R. J. The evolutionary history and spectral tuning of vertebrate visual opsins. *Developmental Biology* **493**, 40–66 (2023).
133. Scully, T. D., Pickett, C. J., Gort-Freitas, N. A., Davidson, B. & Klein, A. M. Divergence of immune cell types in chordate blood. *Current Biology* **35**, 5938-5954.e12 (2025).
134. Lemaire, L. A., Cao, C., Yoon, P. H., Long, J. & Levine, M. The hypothalamus predates the origin of vertebrates. *Sci. Adv.* **7**, eabf7452 (2021).
135. Reilly, M. B., Cros, C., Varol, E., Yemini, E. & Hobert, O. Unique homeobox codes delineate all the neuron classes of C. elegans. *Nature* **584**, 595–601 (2020).
136. Simakov, O. & Wagner, G. P. The application of irreversible genomic states to define and trace ancient cell type homologies. *Evodevo* **16**, 5 (2025).
137. Schultz, D. T. & Simakov, O. Topological Approaches in Animal Comparative Genomics. *Annual Review of Animal Biosciences* **14**, 17–48 (2026).
138. Tertyshnyk, Y. et al. Genome expansions and regulatory contact entanglement help preserve ancestral metazoan synteny. Preprint at https://doi.org/10.64898/2026.04.14.717158 (2026).
139. Scala, F. et al. Phenotypic variation of transcriptomic cell types in mouse motor cortex. *Nature* **598**, 144–150 (2021).
140. Briscoe, S. D. & Ragsdale, C. W. Homology, neocortex, and the evolution of developmental mechanisms. *Science* **362**, 190–193 (2018).