The origin and evolution of cell types

Detlev Arendt[1,2], Jacob M. Musser[1]*, Clare V. H. Baker[3], Aviv Bergman[4,5], Connie Cepko[6], Douglas H. Erwin[5,7], Mihaela Pavlicev[8], Gerhard Schlosser[9], Stefanie Widder[10], Manfred D. Laubichler[5,11,12] and Günter P. Wagner[13–15]*

Abstract | Cell types are the basic building blocks of multicellular organisms and are extensively diversified in animals. Despite recent advances in characterizing cell types, classification schemes remain ambiguous. We propose an evolutionary definition of a cell type that allows cell types to be delineated and compared within and between species. Key to cell type identity are evolutionary changes in the 'core regulatory complex' (CoRC) of transcription factors, that make emergent sister cell types distinct, enable their independent evolution and regulate cell type-specific traits termed apomeres. We discuss the distinction between developmental and evolutionary lineages, and present a roadmap for future research.

Cellular modules
Protein complexes, pathways and molecular machines that make up cell structure and function.

Cell type homology
Cell types that trace back to the same cell type in a common ancestor.

Cell phenotypic convergence
Cell types that are phenotypically similar due to independent changes occurring in separate evolutionary lineages.

Concerted evolution
Similar phenotypic changes that occur simultaneously across different cell types of the same species as a result of altering genetic information shared among the cell types.

Sister cell types
Cell types arising by the splitting of an ancestral cell type into two descendant cells via the process of individuation.

Correspondence to D.A., M.D.L and G.P.W.
arendt@embl.de; manfred.laubichler@asu.edu; gunter.wagner@yale.edu

doi:10.1038/nrg.2016.127
Published online 7 Nov 2016

For centuries, cell types have been defined phenotypically by their structure and function. Also referred to as morphotypes[1], cell types are named according to obvious characteristics of their overall morphology (for example, rod, cone, hair cell and ganglion cell) or to the person who discovered them (for example, Merkel cell). The phenotypic identification of cell types has been refined using modern imaging techniques and functional characterization[2]. For example, retinal ganglion cells have been classified based on electrophysiological recordings, which correlate well with morphology or dendritic stratification patterns. Other studies have used gene expression profiling, fuelled by the rapid progress in single cell RNA sequencing (RNA-seq) and other high-throughput approaches[3–5]. In addition, the modular composition of cell types has come into focus. Typically, cellular functions require the cooperation of many proteins and other biomolecules that constitute small molecular machines, which we refer to as cellular modules[6–10]. Consequently, a cell type has generally been envisaged as expressing an assemblage of cellular modules that perform discrete subfunctions. However, several major problems have rendered current cell type classifications ambiguous. These problems relate to the fact that cell types are the product of an evolutionary process that has shaped their diversity[11].

First, prevailing cell type classification schemes commonly categorize cells according to morphological or molecular similarity. However, from an evolutionary perspective, similarity may arise for different reasons, not all of which reflect *bona fide* cell type homology due to inheritance from a common precursor (FIG. 1). For instance, different cell types can evolve similarity by cell phenotypic convergence. Alternatively, cell types might resemble each other because of concerted evolution[12], which occurs through mutations affecting genetic information used by multiple cell types. It is obvious that these processes have different impacts on morphological and molecular similarity measures; thus, these processes must be distinguished to achieve a clear understanding of cell type identity.

Second, the number of cell types has changed during animal evolution. For instance, basal metazoans have relatively few cell types, indicating that there was a large expansion of cell type diversity before the bilaterian ancestor[13]. New cell types have also appeared, or been lost, in many extant animal clades, including vertebrates[14]. As cell types are inherited through the genome, they must be rebuilt each generation and necessarily share some common developmental history. Thus, gaining knowledge of the changes in cell type number that have occurred through modifications to the genome during evolution and how these relate to developmental lineage is essential to understanding cell types.

In this Review, we propose a new evolutionary definition for cell type that addresses the above issues. We then review our current understanding of the molecular mechanisms underlying cell type identity and discuss examples of the birth of sister cell types. Subsequently, we outline how new cell type-specific phenotypic features evolve. We also disentangle cell type identity from developmental lineage and illustrate how the evolutionary and developmental histories of a cell type often differ. Finally, we present a roadmap for future cell type research, which will be facilitated and guided by the evolutionary viewpoint on cell type identity.

An evolutionary definition of cell type

Here, we define a cell type as 'a set of cells in an organism that change in evolution together, partially independent of other cells, and are evolutionarily more closely related

Author addresses

[1]Developmental Biology Unit, European Molecular Biology Laboratory, Meyerhofstrasse 1, 69012 Heidelberg, Germany.
[2]Centre for Organismal Studies, University of Heidelberg, Im Neuenheimer Feld 230, 69120 Heidelberg, Germany.
[3]Department of Physiology, Development and Neuroscience, University of Cambridge, Anatomy Building, Downing Street, Cambridge CB2 3DY, UK.
[4]Department of Systems and Computational Biology, Dominick P. Purpura Department of Neuroscience, and Department of Pathology, Albert Einstein College of Medicine, 1301 Morris Park Avenue, Bronx, New York 10461, USA.
[5]Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, New Mexico 87501, USA.
[6]Department of Genetics and Department of Ophthalmology, Howard Hughes Medical Institute, Harvard Medical School, Boston, Massachusetts 02115, USA.
[7]Department of Paleobiology, MRC-121, Smithsonian Institution, PO Box 37012, Washington, DC 20013–7012, USA.
[8]Cincinnati Children's Hospital Medical Center, Perinatal Institute, Cincinnati, Ohio 45229–3026, USA.
[9]School of Natural Science and Regenerative Medicine Institute (REMEDI), Biomedical Science Building, Newcastle Road, Galway, Ireland.
[10]CUBE, Department of Microbial Ecology and Ecosystem Sciences, University of Vienna, Althanstrasse 14, 1090 Vienna, Austria.
[11]School of Life Sciences, Arizona State University, PO Box 4501, Tempe, Arizona 85287–4501, USA.
[12]Marine Biological Laboratory, 7 MBL Street, Woods Hole, Massachusetts 02543, USA.
[13]Yale Systems Biology Institute and Department of Ecology and Evolutionary Biology, Yale University, New Haven, Connecticut 06511, USA.
[14]Department of Obstetrics, Gynecology and Reproductive Science, Yale University Medical School, New Haven, Connecticut 06510, USA.
[15]Department of Obstetrics and Gynecology, Wayne State University, Detroit, Michigan 48201, USA.
*These authors contributed equally to this work.

to each other than to other cells'. That is, cell types are evolutionary units with the potential for independent evolutionary change. For cell types, being an evolutionary unit necessarily implies that some genomic information exists that is used only by the cells of a given type and not by other cells[12,15–17]. Only then does a cell type have the potential to undergo evolutionary changes that do not affect other cells. The specific nature of this cell type-specific genomic information — both coding and non-coding — determines cell type identity.

The increase in cell type-specific genomic information that is required for the generation and phenotypic specialization of a new cell type is referred to as genetic individuation[17,18]. Genetic individuation starts the evolutionary subdivision of an initially homogenous set of cells (the ancestral cell type) into descendent sister cell types[19–21]. At the heart of this process are genomic changes that enable incipient sister cell types to express and maintain distinct gene expression programmes. This capacity of a newly born cell type to express genes or combinations of genes that are not expressed by other cells requires the cooperation of a unique combination of transcription factors. We therefore propose that the formation of a new cell type identity requires the evolution of a unique cell type regulatory signature that includes a cell type-specific core regulatory complex (CoRC). This CoRC comprises the set of transcription factors and their cooperative interactions that first enabled the evolution of independent gene expression, and thereby made the new cell type distinct from its evolutionary sister cell type.

Our new cell type definition recognizes and emphasizes the important distinction between cell type identity and the specific phenotype of a cell type in a particular species. That is, cell type identity is defined by the regulatory mechanisms that enable and maintain the distinct gene expression programme of a cell type within the organism. A homologous cell type remains recognizable across species, even when it acquires lineage-specific phenotypic differences, owing to strong evolutionary conservation of these regulatory mechanisms[17]. Conversely, the presence of distinct CoRCs helps to identify cell types that have evolved similar morphology or function through evolutionary convergence, for example the striated muscles of vertebrates and *Drosophila melanogaster*[22]. Below, we elaborate on the empirical basis for this concept as well as its consequences for future research.

Molecular basis of cell type identity

All cells of a multicellular organism can, in principle, access the same genomic information. However, gene expression regulation restricts specific parts of that information to subsets of cells. A growing body of evidence from diverse cell types suggests that the control of cell type identity in differentiated cells is based on a flat hierarchy of gene regulation. That is, a small set of transcription factor genes directly control the majority of cell type-specific effector genes[23–28] and mediate the distinct response of a cell type to common signals[29–31]. The flat architecture of regulation of differentiated cells differs from that found in developmental gene regulatory networks. Gene regulatory networks often exhibit greater hierarchy[32], reflecting spatiotemporal coordination of different developmental events. The set of transcription factors that control cell type-specific gene expression in differentiated cells have been given various names[23–25], and we refer to them here as terminal selectors (as defined by Hobert[24]) (FIG. 2A). The role of terminal selectors in establishing and maintaining postmitotic cell identity has been extensively documented for the nervous system[33]. The importance of terminal selectors in controlling cell type identity is supported by the observation that the forced expression of one or a few of these factors is sufficient to alter cell type identity (reviewed in REF. 23). One classical example is the transformation of fibroblasts into skeletal muscle cells by the forced expression of the transcription factor myoblast determination protein (MYOD)[34]. Follow-up work revealed that this switch in identity is facilitated by autoregulatory feedback loops, in which terminal selectors positively regulate their own expression[23,24,26,33]. This example explains why the forced expression of one selector gene can lead to the activation of the entire set of selector genes and their downstream target genes that are typical for a cell type. The striking capacity of terminal selectors to transform cellular identities in vertebrates and invertebrates has recently been compared to homeotic transformations at the level of whole body parts[35].

There is growing appreciation that terminal selectors are functionally cooperative at the level of protein–protein interactions[36]. One example is the specification

Evolutionary units
Modular biological entities capable of evolving as a cohesive unit and at least partially independently of others (for example, genes, cell types and species).

Genetic individuation
The evolutionary independence of cell types resulting from the differential use of genomic information.

Core regulatory complex
(CoRC). A protein complex composed of terminal selector transcription factors that enables and maintains the distinct gene expression programme of a cell.

Terminal selectors
A set of transcription factors that directly regulates the cell type-specific set of effector genes and represses alternative cell type identities.

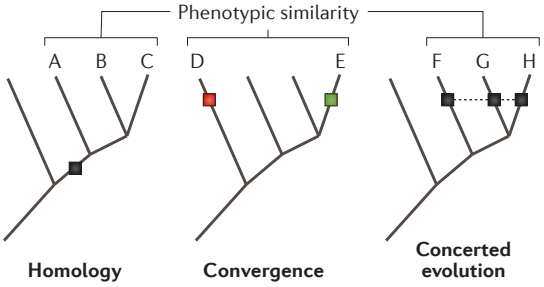


Figure 1 | **Evolution of phenotypic similarity.** Different patterns of evolutionary change can result in cell types A–H in the same organism with similar phenotypes. Cell types A–C share similarity owing to inheritance of a phenotypic change (black box) that occurred in an ancestral cell type. Cell types D and E undergo convergence, arriving at phenotypic similarity independently via two different evolutionary changes (red and green boxes). In cell types F–H, a single evolutionary change in shared genetic information, such as a shared gene or enhancer, results in concerted phenotypic changes across multiple cell types (black boxes).

of the V2a interneuron and the spinal motor neuron in the amniote spinal cord. V2a interneuron identity is conveyed by LIM homeobox 3 (Lhx3) and nuclear LIM receptor (NLI) in the form of a tetrameric complex $(Lhx3:NLI)_2$, whereas motor neuron identity is conferred by Lhx3, NLI and ISL LIM homeobox 1 (Isl1) in the form of a hexameric complex $(Lhx3:Isl1:NLI)_2$ (REF. 37) (FIG. 2B). Similarly, in *Caenorhabditis elegans*, heterodimers of LIM and other homeodomain factors control the identity of various neuronal cell types, such as glutamatergic touch receptors[38]. In the immune system, different lineage decisions and differentiated cell types have also been shown to depend on specific protein–protein interactions[23,39]. For instance, erythroid and myeloid fate is influenced by competition between CREB binding protein (CBP) and PU.1 for binding with the GATA binding protein 1 (GATA1) transcription factor (reviewed in REF. 23). These examples reveal an important principle of cell type identity in that protein–protein interactions within these complexes trigger the suppression of alternative cell fates and the switch-like behaviour of cell fate decisions[36,40–43]. Therefore, cell type-specific gene expression not only requires activity of a specific combination of terminal selectors but also critically depends on their physical cooperativity. It is worthwhile to call this cell type-specific regulatory mechanism by a distinct name — the CoRC [17]. CoRCs are the actual molecular agents that enable differential gene expression among distinct cell types, and it can be argued that CoRCs are even more central to the ability of a cell to realize qualitatively different gene regulatory states than gene regulatory networks[40,43]. Note that in many cases, the CoRC comprises both terminal selectors and more general cofactors such as NLI[44]. Furthermore, terminal selectors may sometimes distribute to more than one CoRC per cell type that, together, co-regulate effector genes and regulate each other, as shown for the complex comprising transcription factor E2-α (TCF3), pancreas transcription factor 1 subunit-α (PTF1A) and recombining binding protein suppressor of hairless (RBPJ) in GABAergic neuron specification[45,46]. The full set of terminal selectors driving differentiation, together with microRNAs[47–51], splicing complexes[52–56] and any other regulatory mechanism that establishes and maintains the differential use of genomic information of a cell within the organism, represents the cell type 'regulatory signature' (as defined by Arendt[19,57]) or 'character identity network' (as defined by Wagner[17]).

Evolution of new cell type identities

The key to the origin of a new cell type is its regulatory independence; that is, the mechanisms for regulating and evolving gene expression independently of other cells. Without a hard-wired regulatory programme that enables cell type-specific gene expression, and unique transcriptional responses to common cellular and environmental signals, there is no stable cell type identity. Thus, the evolution of a new cell type necessarily involves the evolution of a new CoRC, which is the key step for its genetic individuation.

The basic principle of genetic individuation that results in the evolution of sister cell types is illustrated in FIG. 3a,b. A key step is the origin of a new CoRC with at least one new transcription factor (for example, by duplication and divergence of one CoRC component or by recruitment of another factor to the complex) in conjunction with new protein–protein interactions and activities. Subsequently, division of labour events, divergence and neofunctionalization[19] contribute new cell type-specific modules, called apomeres, to the diverging cell type lineages (see below). These become synapomeres if descendant sister cell types undergo additional rounds of evolutionary diversification (FIG. 3a). Most of the coding and non-coding genomic information driving the expression of cellular modules remains shared by both sister cell types; only a fraction becomes specific for each, reflecting incipient individuation (Venn diagram in FIG. 3a,b). One prominent example of a postulated sequence of sister cell type diversification events, accompanied by changes in the CoRCs and apomere formation, is the evolution of ciliary photoreceptor cells[21,58] (FIG. 3c). Note that for these cell types, knowledge of CoRCs is still fragmentary and the full sequence of CoRC and apomere changes remains to be established.

A new CoRC for decidual stromal cells. The emergence of a new cell type and its CoRC has been tracked in much more detail in the evolution of decidual stromal cells (DSCs) and its sister cell type, the neo-endometrial stromal fibroblast (neo-ESF). DSCs differentiate in the uterus of eutherian mammals from neo-ESFs[59,60]. Hierarchical clustering of transcriptomes from various mesenchymal cells showed strong support for a sister cell type relationship between these cell types[61]. Both DSCs and neo-ESFs are evolutionarily derived from an ancestral cell type, called the paleo-ESF, which is unable to differentiate into DSCs. The two derived sister cell types

Apomeres
Derived cell type-specific cellular modules.

Synapomeres
Ancestral apomeres now shared by descendant sister cell types.

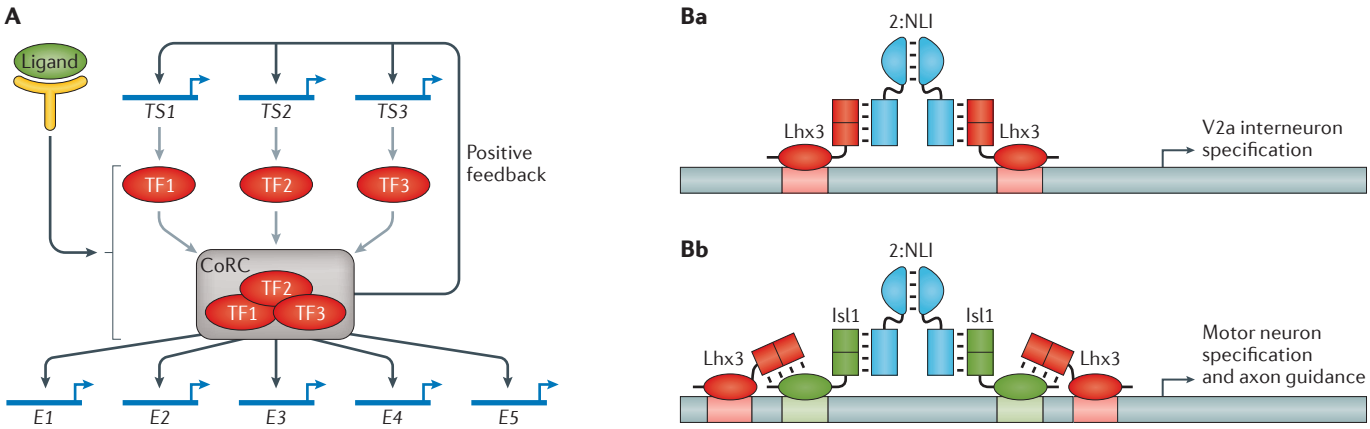
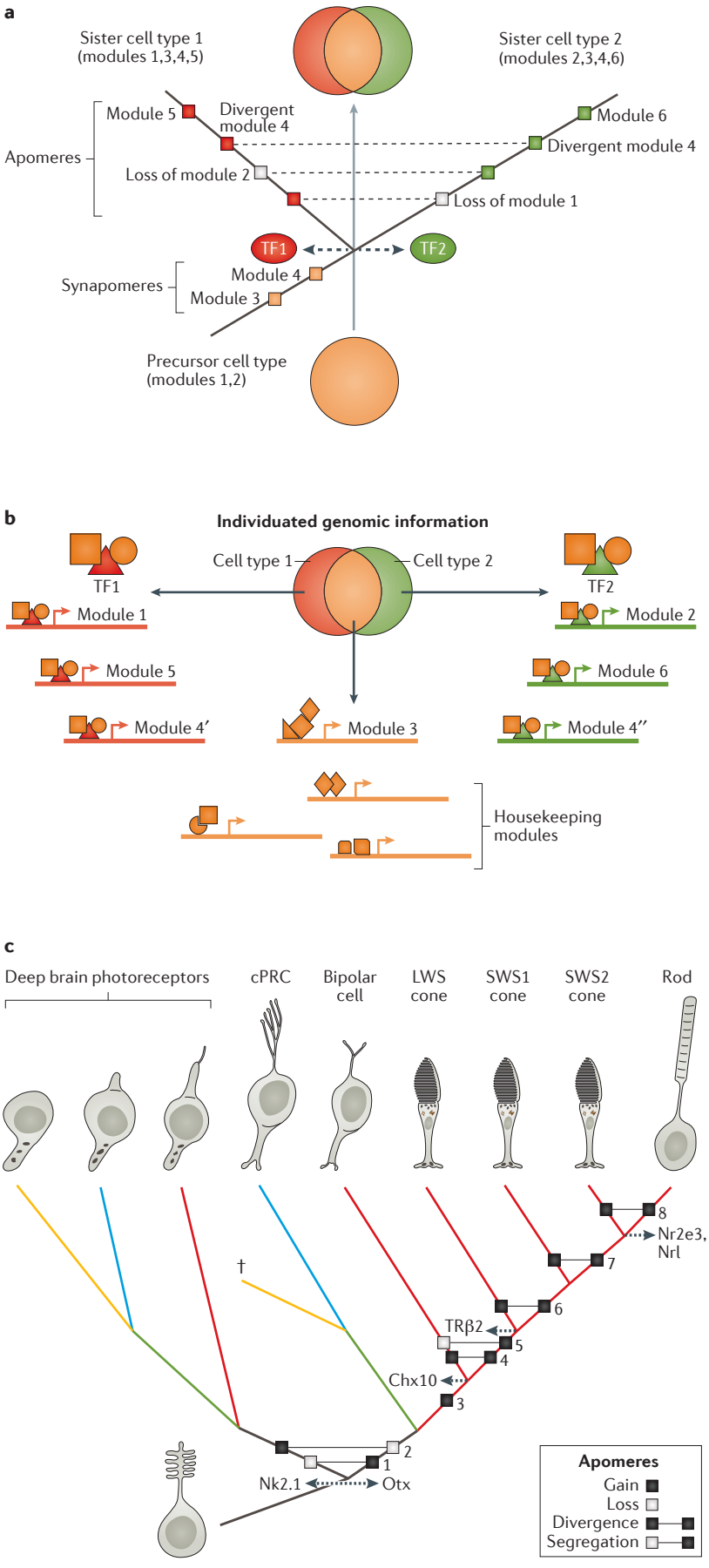


Figure 2 | **The regulatory signature of cell type identity. A** | A model of cell type identity determination. A small set of terminal selector genes (*TS1* to *TS3*) are producing transcription factors (TF1 to TF3), which are modified through the activation of signalling pathways upon binding of a ligand (green oval) and form a core regulatory complex (CoRC). The CoRC is the molecular agent that regulates the downstream effector genes (*E1* to *E5*) and maintains its own expression. In summary, the terminal selector transcription factors cooperatively interact and form a CoRC to regulate cell type-specific gene expression and to enable cell type evolutionary independence. Grey arrows represent translation and complex formation. Black arrows represent regulation. **Ba** | Amniote spinal cord V2a interneurons and motor neurons are sister cell types specified by related CoRCs. V2a interneuron cell type gene expression is regulated by a tetrameric complex composed of two nuclear LIM receptor (NLI) and two LIM homeobox 3 (Lhx3) subunits. **Bb** | In the related motor neuron, ISL LIM homeobox 1 (Isl1) is interposed between NLI and Lhx3 subunits, forming a hexameric complex that regulates gene expression, driving motor neuron specification and axon guidance. Part **B** is adapted with permission from REF. 37, Elsevier.

originated in the stem lineage of eutherian mammals[62], probably to limit the inflammatory reaction triggered by embryo implantation[63]. DSC differentiation requires an entire set of transcription factors: progesterone receptor (PR), forkhead box protein (FOXO1), homeobox protein Hox A10 (HOXA10), HOXA11, HOXD11, HOXD12, GATA2, CCAAT/enhancer-binding protein beta (CEBPB), transcription factor AP-2γ (TFAP2C) and the activating cofactors p300 and CBP[59,61,64–68]. For HOXA11 and CEBPB, there is experimental evidence to indicate that evolutionary changes to their amino acid sequences accompanied the origin of the DSC[69,70] and are now essential for decidual gene expression. Both factors functionally cooperate with FOXO1 via direct protein–protein interactions and phosphorylation[70,71]. For CEBPB, two losses and one gain of phosphorylation sites have made the protein responsive to FOXO1 binding and to glycogen synthase kinase 3 (GSK3) phosphorylation. Notably, comparative data have indicated that physical protein–protein interactions preceded functional cooperation between HOXA11 or FOXO1 and CEBPB[70,72]. Another example of the emergence of novel transcription factor regulatory activity in the context of cell type evolution is the acquisition of neural crest-inducing activity by FOXD3 in vertebrates[73]. In another study, a systematic comparison of transcription factor binding sites between *Drosophila melanogaster* and vertebrates was performed. The results revealed that most factors showed highly conserved specificities; however, those with altered binding appeared to specify novel cell types that form parts of the endocrine and adaptive immune systems[74].

In summary, the evolution of a new regulatory signature is necessary for establishing a novel cell type, which represents a new, individuated evolutionary unit within the organism. In the examples discussed above, this process accompanies an increase in cell type number via sister cell type diversification. Other scenarios are also plausible. In an extreme case, one might envisage the opposite trend, a decrease in cell type number, such as occurred in myxozoans, which are small parasitic cnidarians with few distinct cell types[75]. A decrease in cell type number might occur due to loss of function, such as photoreceptors in cave animals, or via the fusion of cellular identities through the co-activation of CoRCs from previously different cell types, resulting in a functionally hybrid cell[12,76].

The same cell type regardless of form and function. Before moving on, we want to point out that the model of cell type identity discussed here has important implications for our understanding of cell type homology. Our model implies that the independent control of gene expression, which we link to cell type identity, may be mechanistically dissociated from cell phenotype. An illustrative example is bilaterian visceral muscles. In both vertebrates and the annelid *Platynereis dumerilii*, visceral muscles are composed of smooth myocytes that express homologous transcription factors, including myocyte-specific enhancer factor 2 (Mef2), Foxf1 and Nkx3.2 (REF. 22). Visceral muscles in *D. melanogaster* also express these transcription factors, yet have a striated phenotype, suggesting that visceral myocytes co-opted a striated module in an ecdysozoan ancestor[22]. Despite



this phenotypic change, our model recognizes bilaterian visceral myocytes as homologous because they apparently descend from myocytes in the bilaterian ancestor that used a similar regulatory signature. This example shows that the CoRC from homologous cell types is conserved, whereas the phenotype of the cell is more flexible. Hence, this model is the first to explain mechanistically what is meant by the 'same cell type regardless of form and function', which is of course a paraphrase of the classical definition of homology put forward by Owen in 1848 (REF. 77).

Concerted evolution

Relatively few genes may be uniquely expressed in a specific cell type, which has important implications for cell type evolutionary interdependence. FIGURE 3b illustrates that many genes are expressed across multiple cell types. With regard to nervous system evolution, good examples of such genes are the pan-neuronal genes expressed in all or most types of neurons. Other

Figure 3 | **Sister cell types evolve by individuation.** **a** | A model of sister cell type diversification. The starting point was a precursor cell type with two modules (1, 2). Before diversification, two new modules had arisen (3, 4) that became synapomeres after the split. A key step in the formation of sister cell types was the evolution of two distinct core regulatory complexes (CoRCs) employing transcription factors TF1 and TF2. Phenotypically, cell type diversification involved division of labour events (modules 1, 2), module divergence (module 4) and the acquisition of new modules (5, 6). Corresponding modules in the sister cells are connected by dashed lines. **b** | Venn diagram showing shared and cell type-specific genomic information of sister cell types. In this example, cell types 1 and 2 gain evolutionary independence through the duplication of an ancestral transcription factor, resulting in related trimeric CoRCs. The cell type-specific genetic information (red and green) allows cell types to evolve with partial independence. Shared genetic information (orange), including housekeeping modules and the *cis*-regulatory elements driving their expression, leads to concerted evolution of the two cell types when altered. 4′ and 4″ represent derivatives of module 4 that have arisen through module divergence. **c** | Evolution of ciliary photoreceptors (cPRCs) in Bilateria. Different colours indicate phylogenetic cell type splits, with red representing vertebrates, green representing protostomes, yellow representing ecdysozoans, and blue representing lophotrochozoans. Dashed arrows indicate known transcription factor changes[58]. Different apomeres are indicated by numbers: 1, surface-extended cilium; 2, control of reproduction via neurosecretion; 3, visual function; 4, axonal projection or interneuron function; 5, light sensitivity via c-opsin; 6, deployment of the vertebrate long wavelength-sensitive (LWS) or of the vertebrate short wavelength-sensitive (SWS)1–2–rhodopsin duplicate; 7, deployment of the SWS1 or of the SWS27–rhodopsin duplicate; 8, deployment of the SWS2 or of the rhodopsin duplicate. Cross denotes cell type loss. Chx10, ceh10 homeobox-containing homologue; Nr2e3, nuclear receptor subfamily 2 group E member 3; Nrl, neural retina-specific leucine zipper protein; TRβ2, thyroid hormone receptor β2. Part **c** is adapted from REF. 19, Nature Publishing Group.

examples include the contractile genes shared across different muscle cell types[22,78]. Mutations in both coding and *cis*-regulatory sequences of these genes may simultaneously influence several cell types in similar ways, resulting in their concerted evolution[12] (exemplified and explained for a group of sister cell types in FIG. 1). For instance, effector genes are often used by different cell types for similar functions. Mutations that alter protein function will then affect all cell types that utilize that particular protein function, resulting in their concerted evolution. Similarly, alterations to shared regulatory mechanisms of gene expression, such as the loss of an enhancer used by multiple cell types, may cause concerted gene expression evolution in those cell types. It is evident that a large fraction of such shared genomic information dramatically reduces the capacity of cell types to evolve independently and, therefore, their degree of individuation[15].

A prerequisite for concerted gene expression evolution is the sharing of regulatory information by different cell types. A systematic investigation of pan-neuronal gene expression in the nematode *C. elegans* identified enhancers that drive expression across different, overlapping sets of neuronal cell types[79]. Partial deletion of these enhancers resulted in changes to pan-neuronal gene expression in different cell types, indicating the potential for concerted gene expression evolution. However, that study also revealed that cell type-specific terminal selectors also regulate pan-neuronal gene expression[79], allowing individual cell types to independently tune their expression.

Apomere evolution

In addition to changes in CoRCs, genetic individuation involves the emergence of new cellular modules in the incipient sister cell types. For instance, novel cellular modules may emerge, adding new functionality to the cell. In other cases, pre-existing modules may diverge, producing new functional variants, or end up in different sister cell types, triggering division of labour[19]. We refer to new cell type-specific modules, or new variants of modules, as apomeres (analogous to apomorphies at the species level). Similar to the evolution of new CoRCs, apomere evolution typically involves the addition of new proteins or changes in protein structure, function and/or new protein–protein interactions. In contrast to CoRCs, which allow sister cell types to evolve and maintain differences in gene expression, apomeres manifest their distinct cellular phenotypes. Here, we discuss two prevalent modes by which apomeres emerge and evolve: module integration and module divergence.

Module integration. Module integration leads to the emergence of apomeres by combining existing cellular machinery into a new module with emergent functions. Two aspects are key to this process: the evolution of new protein–protein interactions and the coordinated expression of the integrated proteins. One illustrative example is the evolution of the neuronal synapse (FIG. 4). Synapses evolved early in animal evolution; they are not found in the closest relatives of animals, the choanoflagellates[80] or in two early diverging animal lineages[81,82], sponges[83–88] and placozoans[89]. Synapses are highly specialized cell junctions between presynaptic and postsynaptic cells that combine a nearly complete set of adherens junction proteins (and other adhesion-related complexes) with other distinct modules. In the presynaptic terminal, active zone and exocytosis machinery function to release and regenerate neurotransmitter vesicles. In the postsynaptic terminal, neurotransmitter receptors, ion channels, signalling and scaffolding proteins convert the neurotransmitter signal into changes in membrane potential. As synapses often connect two different cell types, presynaptic and postsynaptic terminals represent two distinct cell type apomeres.

Comparative evidence has revealed that many components of the presynapse and postsynapse, as well as junctional machinery, evolved before the origin of synapses, with independent roles. For instance, many postsynaptic density genes are present in choanoflagellates[90], and are co-expressed, and probably interact, in the sponge *Amphimedon queenslandica*[91]. The exocytosis SNARE machinery evolved early in eukaryote phylogeny[92], and orthologues of neuronal SNARE proteins form a complex in choanoflagellates[93]. Finally, ultrastructural evidence has revealed the presence of adherens junctions in sponges[94], which lack synapses, and some evidence has indicated that adherens junctions may even pre-date Metazoa[95–97].

These findings suggest that the presynapse and postsynapse evolved through the integration of exocytosis and receptor machinery alongside pre-existing cell junctions, most likely adherens junctions. Supporting this notion, synaptic adhesion protein complexes have been found to interact with (and regulate) both presynaptic and postsynaptic functions, beyond their conventional junctional role[98–103]. An example is the N-cadherin–β-catenin complex, which regulates vesicle reuptake and short-term plasticity of the presynapse[98,99]. In addition, a recent study investigating co-regulation of synaptic genes during development found correlated expression changes of synaptic homologues in diverse animal species with synapses, but not in the synapse-less sponge *A. queenslandica*[104].

Module divergence. Module divergence occurs when the same cellular module evolves cell type-specific subunits and specialized functions in different cell types (FIG. 4). The widespread occurrence of module divergence was documented in a recent study[105], which investigated stoichiometric variation in 182 well-characterized protein complexes in human and mouse cell types. Notably, more than half of these well-characterized complexes varied in their subunit composition across cell types[105]. Variable protein subunits were enriched for paralogues, with different cell types making use of different paralogous subunits (paralogue switching)[105]. For instance, ADP-ribosylation factor GTPase-activating proteins (ARFGAPs) are important regulators of coat protein I (COPI) vesicular transport between the Golgi complex and the endoplasmic

Pan-neuronal genes
Genes expressed broadly, but not exclusively, in neurons; for example, synaptic and vesicular genes.

Module integration
Evolution of a new functional complex or pathway by colocalization and integration of pre-existing functional machinery.

Module divergence
Evolution of cell type-specific variation in protein complexes or pathways; commonly occurs by gene duplication and divergence.

SNARE
Soluble NSF (*N*-ethylmaleimide-sensitive factor) attachment protein (SNAP) receptor.

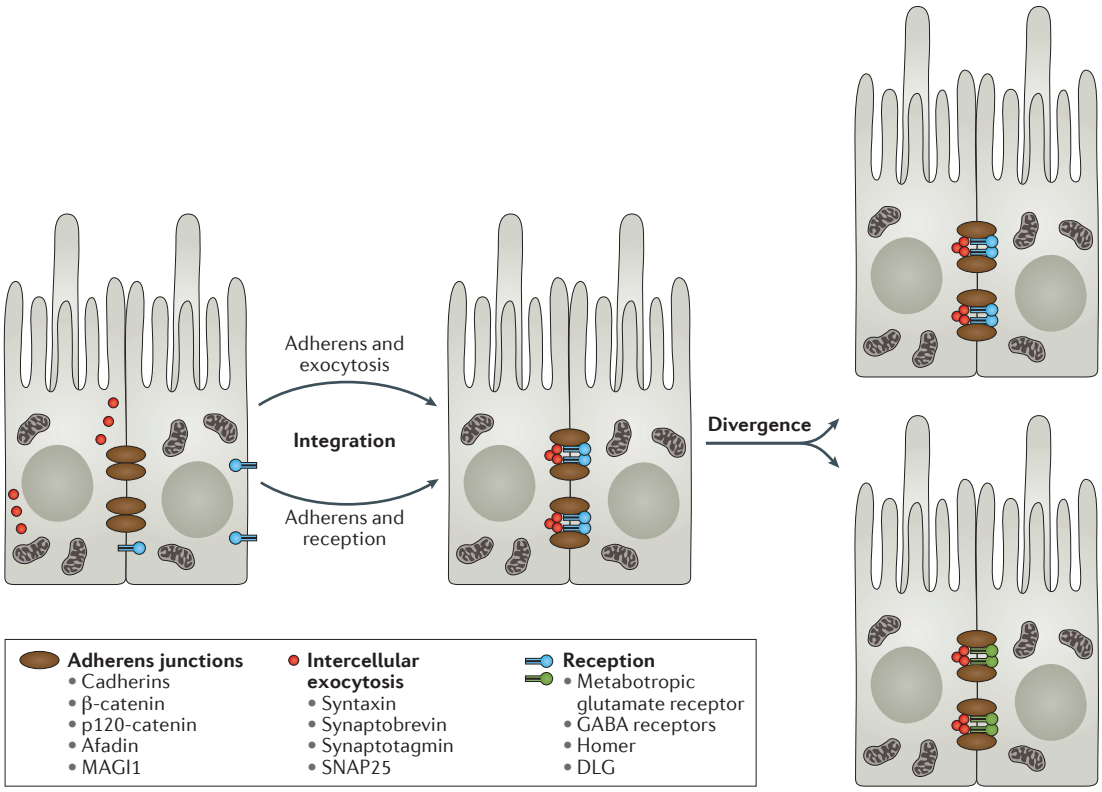


Figure 4 | **Evolution of pre- and postsynaptic apomeres by module integration and divergence.** Neuronal synapses are composed of presynaptic and postsynaptic terminals, united around adherens-like junctions. Evidence from close relatives of metazoa and basal metazoans indicates that most synaptic functional machinery, such as neurotransmitter vesicles (red circles) and receptors (blue bulbs), pre-dates the evolution of a functional synapse. The first proto-synapse may have evolved in early hair cell-like sensory effectors cells, as shown in the figure, through module integration of this pre-existing neurotransmitter machinery around adherens junctions (brown ovals), resulting in pre- and postsynaptic apomeres, and allowing for rapid, directed communication. Subsequent module divergence, for instance through the gene duplication of neurotransmitter receptors (blue and green bulbs), has facilitated the evolution of numerous cell type-specific pre- and postsynaptic apomeres in animal nervous systems. DLG, Discs large; MAGI1, membrane-associated guanylate kinase, WW and PDZ domain-containing protein 1; SNAP25, synaptosomal-associated protein 25.

reticulum. The paralogues ARFGAP2 and ARFGAP3 have diverged functionally[106,107], producing distinct COPI transport apomeres.

Notably, the extent to which modules evolve cell type-specific differences varies substantially depending on their function[105]. Protein complexes involved in energy metabolism, or other housekeeping functions, are generally invariant across cell types. By contrast, protein complexes associated with the regulation of chromatin or subcellular transport have a relatively high proportion of subunits that vary across cell types. The module divergence of transport machinery suggests that one important aspect of cell type evolution is the differential subcellular distribution of proteins. This finding may help account for the numerous morphological differences found among cell types.

Beyond housekeeping modules, evidence for divergence can be found in cellular modules involved in a wide variety of cellular functions. In the synapse, module divergence has occurred extensively[108–111]. For instance, duplication of genes encoding subunits of the postsynaptic scaffold and the glutamate receptor has probably facilitated functional diversification of neurons in the vertebrate brain[109,110]. In another case, individuation of neo-ESFs and DSCs in uterus evolution was accompanied by changes in innate immunity pathways. Neo-ESFs of eutherian mammals lost components of a pro-inflammatory module, including the lipopolysaccharide (LPS) receptor, as well as cofactor CD14, which connects the LPS receptor to Toll-like receptor 4 (REF. 112). This finding suggests that the cell type-specific loss of module subunits can also lead to module divergence. Finally, duplication of the gene encoding fascin (*FSCN*) resulted in the evolution of a new cytoskeletal bundling apomere in the vertebrate retina. FSCN1 is widely expressed across different cell types and regulates the assembly of actin filaments via interactions with protein kinase Cγ (PKCγ) and RAC1 (REF. 113). The paralogue FSCN2 also regulates actin, but is found exclusively in the inner segment of photoreceptor cells and has been implicated in macular degeneration and retinitis pigmentosa[114].

In some cases, a cellular module may diverge such that it is nearly unique, allowing sister cell types maximum freedom to evolve specialized functions. For

example, vertebrate rod and cone phototransduction cascades evolved from a single ancestral cascade (see divergence of function in FIG. 3a). The two rounds of genome duplication during vertebrate evolution[115] led to the duplication of almost all subunits in the ancestral pathway. This event facilitated the evolution of distinct rod and cone phototransduction apomeres[58,116,117], which are responsible for the physiological differences between rods and cones[58,118–122]. For instance, rods are highly sensitive to light but inactivate slowly. By contrast, cones are less sensitive to light but recover rapidly. When the rod paralogue of transducin alpha, an important pathway subunit, was replaced with the cone paralogue, the rods of these mutant mice had decreased sensitivity to light and recovered more rapidly, approaching the physiological characteristics of wild-type cones[119].

Development versus evolution

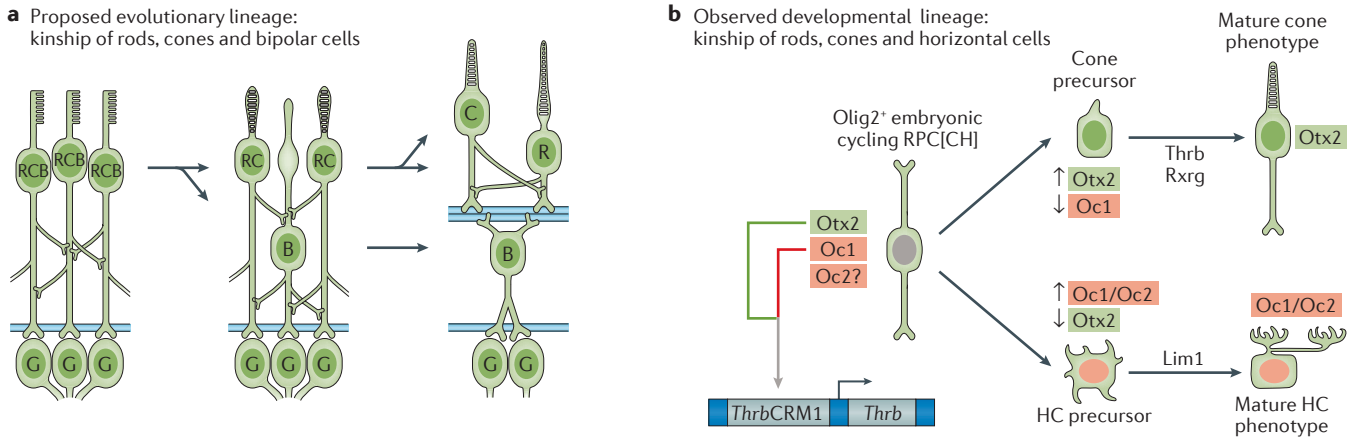
In every generation, all the different cell types of a given organism are built anew through the processes of development and differentiation. It is important to stress that the developmental and the evolutionary lineage of cell types are not necessarily the same. The developmental lineage is represented by the patterns of cell division and fate decisions that ultimately deploy a differentiated cell type at a specific place and time within the organism. By contrast, cell type evolutionary history reflects the diversification of sister cell types through the evolution of new genetic programmes, from the few cell types of metazoan ancestors to the many hundred cell types of most extant bilaterians. These hierarchical trees, representing cell type evolutionary and developmental lineage histories, may be incongruent.

Box 1 | The evolutionary and developmental lineage of retinal cell types

The vertebrate retina is one of the most thoroughly investigated tissues, combining molecular mechanistic detail[125,138] with cellular resolution[139,140]. Therefore, the retina represents an ideal case study for the comparison of developmental and evolutionary lineage.

Cell type homologies within and across species have been hypothesized based on cellular morphologies and molecular characteristics[21,58]. For instance, kinship of ciliary photoreceptor cells was proposed for vertebrates, ascidians, amphioxus and possibly annelids on the basis of ciliary morphology and shared terminal selectors, such as the homeobox proteins Rx and Otx[141]. Also, a sister cell type relationship of ciliary photoreceptors in the retina, the pineal gland and the hypothalamus was suggested and, importantly, extended to the bipolar cells of the vertebrate retina[21,58] (see the figure, part **a**). The bipolar cells are interpreted as direct sisters to the ciliary photoreceptors that arose by division of labour, with the rods and cones 'inheriting' the photoreceptive properties, and the bipolar cells inheriting the axonal projection properties. Again, this hypothesis is supported by similar terminal selectors, containing Otx2, Crx and Rx, and possible synapomeres, such as ribbon synapses. In stark contrast, the retinal ganglion (and possibly, amacrine) cells appear to be linked to non-retinal mechanoreceptor cells such as hair cells (BOX 2), based on similar terminal selectors (Atonal homologue 1 (Atoh1) or Atoh5 and Brn3). Finally, the GABAergic horizontal cells show a divergent core regulatory complex (CoRC) (pancreas transcription factor 1 subunit-α (Ptf1a) and Onecut (Oc)) that sets them apart from the other lineages.

The developmental lineage has similarly been investigated in detail in mice[142] and fish[143]. One well-investigated example case is the origin of cone photoreceptors and horizontal cells from the same precursor that co-expresses Otx2 and Oc[125] (see the figure, part **b**). One obvious and intuitive interpretation of this shared developmental lineage would be an evolutionary sister cell type relationship that, however, would be at odds with the proposed closer link between rods and cones and bipolar cells. The serial sister cell type concept may help us to understand this seeming contradiction. This concept would interpret the developmental lineages in the retina as serial duplicates of an ancient stem cell-like system that produced cell types specified by Atonal-, Ptf1a- and ASC-related basic helix–loop–helix (bHLH) factors. These cell types would have evolved into ganglion or amacrine (Atonal-related), horizontal (Ptf1a-related) and, possibly, bipolar or photoreceptor (ASC-related) cells. This finding would imply that retinal cell types may have serial sister cell types outside the retina, including, for example, the following: Atoh1-positive, Brn3-positive mechanosensors (BOX 2); Ptf1a-positive GABAergic cells in cochlear nucleus[144], cerebellum[145] and alar plate[46]; and other ASC-positive ciliary-type sensory cell types of the apical nervous system[146]. Part **a** is from REF. 19, Nature Publishing Group. Part **b** is reproduced with permission from REF. 125, Cell Press/Elsevier.



B, bipolar cell; C, cone cell; CRM, *cis*-regulatory module; G, ganglion cell; HC, horizontal cell; Lim1, homeobox protein Lim-1 (also known as Lhx1); Olig2, oligodendrocyte transcription factor 2; R, rod cell; RC, rod and cone precursor cell; RCB, rod, cone and bipolar evolutionary precursor cell with both photoreceptor and interneuron functions; RPC, retinal precursor cell; RPC[CH], RPC biased towards the production of cones and horizontal cells; Rxrg, retinoid X receptor-γ; Thrb, thyroid hormone receptor-β.

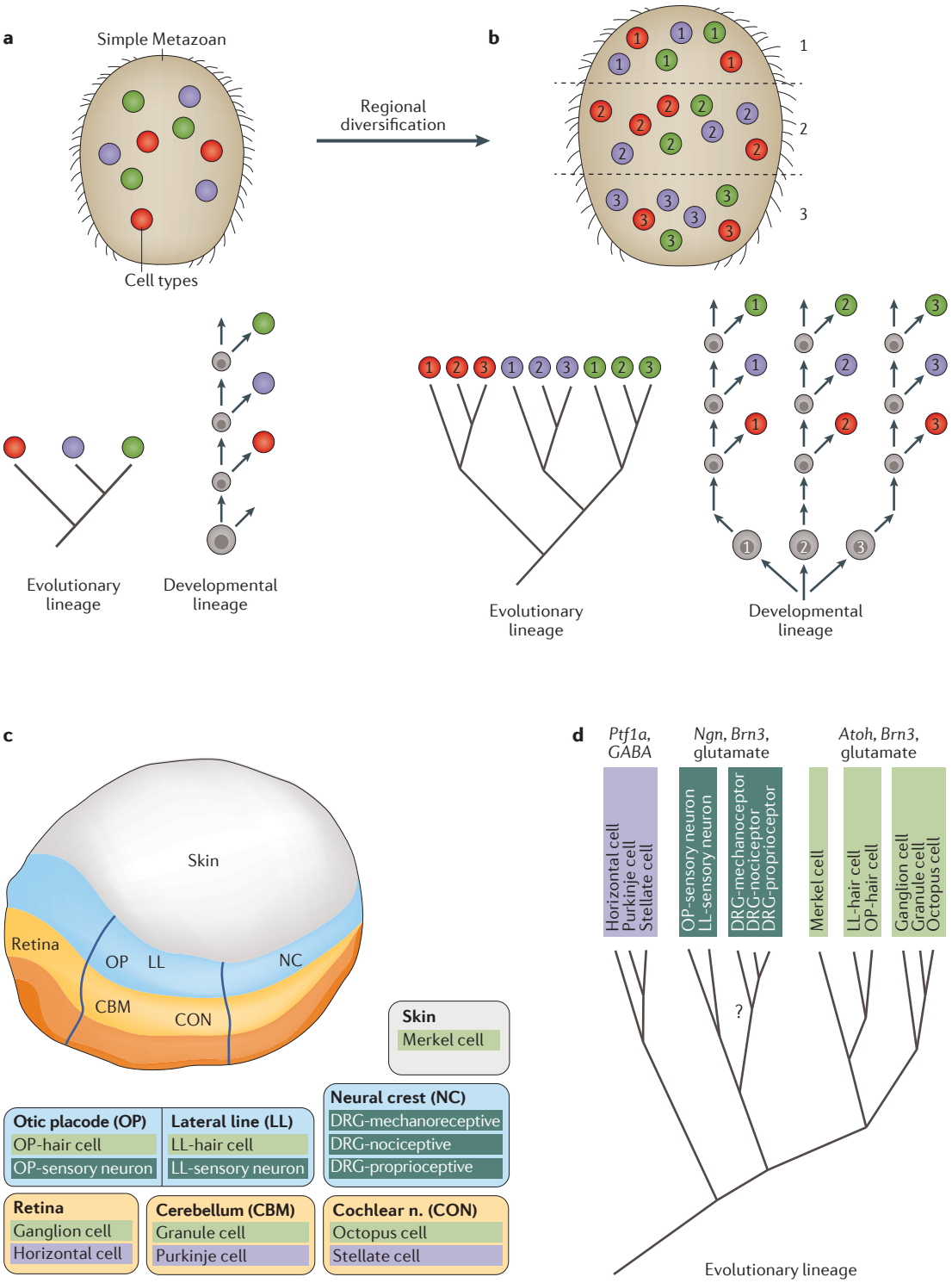
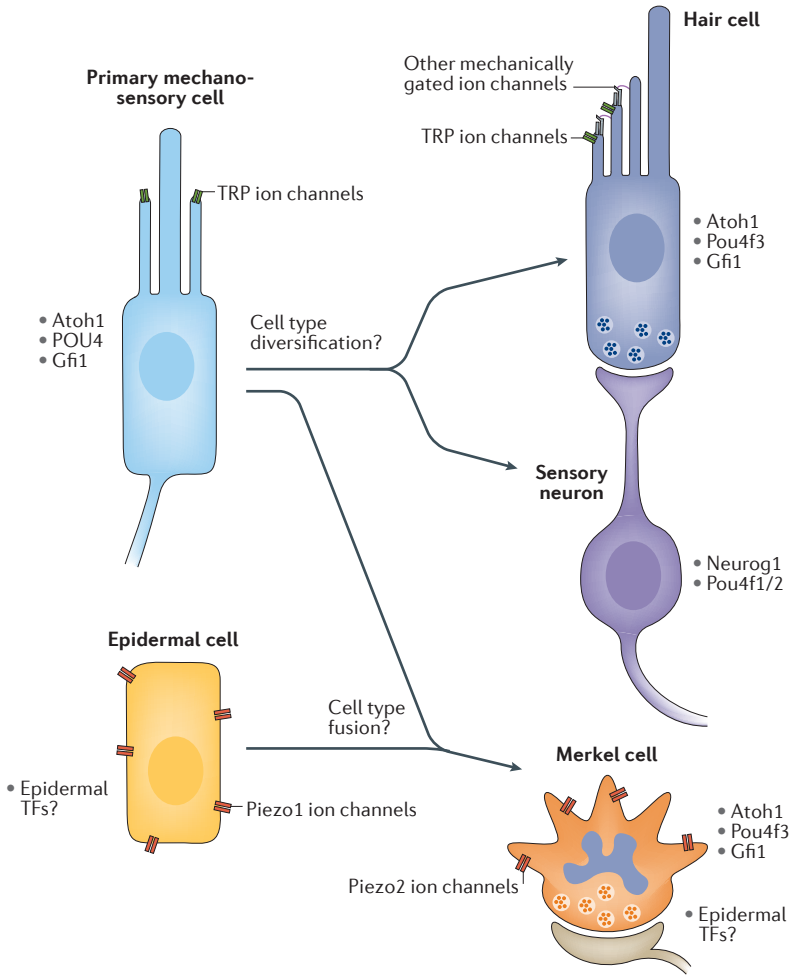


Figure 5 | **Interrelationship of developmental and evolutionary cell type lineages. a** | Ancestral state. Three evolutionarily related cell types are homogenously distributed across the body in a hypothetical simple metazoan, arising from a stem cell-like developmental lineage. **b** | Derived state. Cells have diversified regionally, giving rise to region-specific serial sister cell types. Within a region, cells arise from common stem cells so that developmental and evolutionary lineage differ. **c** | Possible serial sister cell types are shown in the same colour, developing from anteroposterior and mediolateral regions of the vertebrate body. The cell types of each body region are listed in the boxes. Box colours match that of the respective body region. **d** | Evolutionary lineage of the same cell types in a cell type tree, with shared core regulatory complex transcription factors and transmitter indicated for each group of serial sister cell types. *Atoh*, Atonal homologue family genes; cochlear n., cochlear nucleus; DRG, dorsal root ganglion; *Ngn*, neurogenin family genes; *Ptf1a*, pancreas transcription factor 1 subunit-α and related genes.

Box 2 | **Serial sister cell type and cell type fusion in sensory cells**

Vertebrate hair cells (see the figure) are secondary sensory cells (lacking axons) specialized to detect local fluid movement. Inner ear hair cells are used for hearing and balance, whereas lateral line hair cells (in fish and amphibians) sense external water movement. A network of transcription factors (TFs), including Atoh1 (an Atonal-related basic helix–loop–helix (bHLH) factor), Pou4f3 (also known as Brn3c) and growth factor independent protein 1 (Gfi1), are important for hair cell specification in several neurogenic placodes[147–150]. Additionally, orthologous factors specify a subset of mechanosensory, photosensory and chemosensory cells in different body regions in insects and annelids[128,129,151,152], suggesting that such sensory cells represent serial sister cell types. Other mechanoreceptors may have evolved convergently from sensory cells specified by ASC-related bHLH factors. Vertebrate hair cells and their afferent neurons may have arisen as sister cell types by division of labour from primary mechanoreceptors possessing an axon, as found in invertebrates[130]. Supporting this notion, these sensory neurons share common clonal origins with hair cells and are specified by neurogenin 1 (Neurog1), which arose by duplication of an Atonal-like gene in early bilaterians[153,154]. Similarly, lateral line electroreceptors plausibly arose as sister cell types of hair cells, although molecular evidence is lacking[155]. Epidermally derived Merkel cells, another type of vertebrate mechanoreceptor[156,157], are secondary sensory cells that also rely on Atoh1, Pou4f3 and Gfi1 (REFS 149,157,158). Merkel cells may be axon-less homologues of the Atonal-dependent caudal epidermal sensory neurons of ascidians[159,160]. However, Merkel cells use Piezo2, a paralogue of the Piezo1 channel, for mechanotransduction[157]; Piezo1 senses membrane stretching in epidermal cells[161]. Hence, Merkel cells may have evolved through the co-option of a mechanosensory gene regulatory network into epidermal cells (cell type fusion) rather than by divergence of a pre-existing sensory cell type, although more evidence is needed to substantiate either scenario.



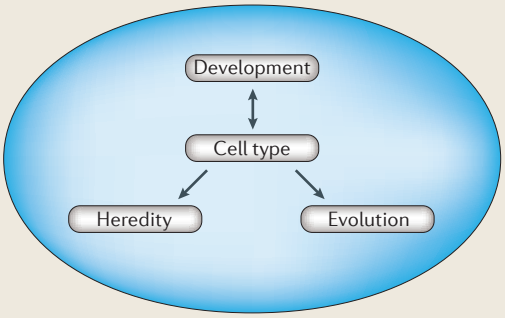
TRP, transient receptor potential.

A disconnect of developmental and evolutionary lineage. A disconnect between cell type identity and developmental lineage was first indicated in late nineteenth century embryological experiments conducted by Vincenzo Colucci. In his research, Colucci showed that the lens in the eyes of urodeles (an order of amphibians) could be regenerated from the iris (reviewed in REF. 123). More examples are found in the nematode *C. elegans*, for which a full developmental lineage is available. Cells belonging to the same cell type can be derived from highly distinct branches of the developmental lineage, whereas a single sublineage can give rise to markedly different cell types, such as muscles and neurons[124]. In line with this notion, other bilaterian stem cell-like developmental lineages also often produce cell types that are markedly distinct in terms of the genomic information they use. For instance, horizontal cells and cone photoreceptors in the vertebrate retina, which are strikingly different both morphologically and molecularly[125], are produced from the same lineage precursors (BOX 1). Also, in the developing nervous system of the fly, neural progenitors generate highly divergent neuronal and glial cell types[126]. And, as in nematodes, this cell type disparity contrasts with the remarkable molecular and phenotypic resemblance of cell types produced by unrelated developmental stem cell-like sublineages. For example, horizontal cells in the vertebrate retina resemble the inhibitory GABAergic interneurons found in the cerebellum, hindbrain nuclei and alar plate of the spinal cord (BOX 1).

Serial sister cell types. How then are we to understand the apparent disconnect between evolutionary and developmental histories? We propose the concept of serial sister cell types as a solution to this conundrum (FIG. 5). First, based on comparative observations of development, it is now well established that early animal evolution involved the repeated subdivision of the animal body into distinct regions. We propose that these regionalization events also led to the duplication and subsequent diversification of at least one of the cell types that populated that region. This process produced an iterated series of topographically separate sister cell types that we refer to as serial sister cell types. It is plausible that these cell type duplication events also led to the evolution of serial sister stem cells, as virtually all animal cell types co-occurring in one region develop from asymmetrically dividing, multipotent stem cell-like cells (FIG. 5a,b). As a result, serial sister cell types arise from different regions and are produced by different stem cells, despite sharing a close evolutionary relationship (compare also the sublineage concept of Sternberg and Horvitz[127]).

We propose that serial sister cell types are widespread across the animal body. In addition to the Ptf1A-positive GABAergic interneurons, Atonal- and Brn3-positive glutamatergic mechanosensory and photosensory neurons are found in different regions of the vertebrate body, including the dorsal neural tube and the sensory placodes. Indeed, the possible homology of some of these cell types has long been recognized[128–130] (FIG. 5c,d). BOX 2 exemplifies the principle of serial sister cell types for the

Box 3 | **Cell types at the intersection of cell, developmental and evolutionary biology**



Cell theory is closely associated with the search for fundamental units of life and their properties. Throughout the nineteenth and early twentieth centuries we saw an emphasis on understanding the basic properties of cells, both structurally and functionally[162–165]. Different cell types were recognized in the context of microscopic anatomy and histology as well as pathology[166,167]. Links to understanding developmental and evolutionary processes emerged in the context of studies of heredity and differentiation (see the figure). Cell lineage studies traced the genealogy of cells within the developing embryo, thus providing the first detailed descriptions of the patterns of differentiation and specialization of cells[168]. The goal here was to establish connections between more differentiated parts of organisms and early embryonic cells and to connect development and heredity. Studies of fertilization had established a close link between chromosomes and heredity, culminating in the Boveri–Sutton chromosomal theory of inheritance[165]. The question then became one of identifying the causal determinants of both heredity and differentiation. Studies conducted by Boveri and others suggested a central role of the chromosomes that soon led to the localization of specific hereditary units or 'genes' for recognizable phenotypic traits at specific regions on the chromosomes by the Morgan group[169]. The mechanisms by which those hereditary units cause phenotypes, including cell types, remain unclear. But Boveri, in his Rektoratsrede of 1906 (REF. 170), offered a conceptual suggestion that can be seen as a direct antecedent to the ideas proposed here: "Even the most complex individual emerges from a simple cell, the egg; in this cell we find the conditions for all traits of the final outcome. One can, therefore, call these complexes of embryonic conditions the fundamental traits of an organism; to analyse those is one of our future tasks. We can already state that this substrate of anlagen cannot be thought of as a mere conglomerate, but rather as a system, one that is all the more complex, the higher the organism. Within this system there must exist constructive mechanisms, ranging from the most general to the specific, which are ordered in a hierarchical fashion and which interact in a lawful way" (translation by M.D.L.).

Atoh-positive hair cells in the vertebrate otic and lateral line placodes. The serial sister cell type concept thus explains the fundamentally different trees of relatedness in development and evolution; developmentally related cell types are not necessarily evolutionarily related and vice versa.

A roadmap for future research

Our evolutionary definition of a cell type will facilitate comparative cell biological and organismal research in a broad range of disciplines, from development to physiology or neurobiology, to name but a few. Whenever a comparison between species is involved, it is pivotal to know whether similar, related or unrelated cells are being compared. For example, the field of comparative connectomics will benefit from a more precise knowledge of the identity of the neuron types that actually constitute the neural circuits found in different species. Different cell types, even those that are functionally similar, may exhibit distinct responses to stimulation[29–31]. Also, the comparison of immune systems across vertebrates, or even across bilaterians, requires a rigorous framework for cellular homologies. The recognition that cell type identity and specific function depend on only a small set of genetic information, including CoRCs and apomeres, enables the efficient identification and categorization of cell types across the animal tree. Furthermore, by disentangling cell type identity from phenotype and developmental lineage, our approach enables a better understanding of the link between development, heredity and evolution (BOX 3). Our effort to overcome the limitations of phenotypic cell type classification schemes is analogous to the conceptual transition from Linnean to evolutionary classification at the species level and should be equally beneficial.

One of the most exciting areas will be comparative cell biology itself, including efforts to track the history of cell types within and across species. The benefit of this research programme is not limited to clarifying historical patterns of tissue diversification but will also help in understanding the changes in gene regulatory network structure in development and evolution. Clearly, tracking cell type evolution will require specialized tools. So far, the comparison of cellular transcriptomes works reasonably well within species[61,131,132] but is considerably more challenging across species[12]. Eventually, we will need a deeper mechanistic understanding of the process of genetic individuation, in much the same way as gene phylogenies have required models of DNA and protein sequence evolution. Moreover, we will need a strategy to integrate proteomic and functional data to infer cell type interrelationships. One exciting outcome of reconstructing cell type evolution will be to infer the exact order of apomere emergence and the molecular details of apomere emergence and diversification, resulting in an evolutionary history of cellular module and organelle evolution in animals.

Finally, many other exciting theoretical and empirical questions can be addressed. Are there generalizable gene regulatory network motifs that are necessary for cell type-specific expression patterns, and how do these motifs and associated networks evolve[133]? In addition, how is gene expression heterogeneity, often uncovered in single-cell transcriptome data[134–136], related to core

Cell type fusion
Co-option of a second core regulatory complex into an existing cell type, creating a cell type hybrid of two different ancestral cell type identities.

Serial sister cell types
Sister cell types that arise from different developmental lineages or regions of the body.

Sensory placodes
Thickened patches of embryonic head ectoderm that contribute sensory receptor cells, secretory cells and supporting cells to peripheral sense organs, and/or sensory neurons to cranial ganglia.

Comparative connectomics
A research programme comparing the neuronal connection networks between species.

regulatory complexes and the ability of cell types to vary in evolution? Another challenging area is the co-analysis of cellular identity and developmental lineage. Our model of sister cell type evolution suggests that there can be discordance between evolutionary and developmental lineage. To understand this discordance will require the careful distinction of cell types across different developmental stages and lineages. Here, capturing CRISPR-edited lineage markers[137] with single cell RNA-seq will add cell type resolution to lineage studies.

1. Valentine, J. W. in *Keywords and Concepts in Evolutionary Developmental Biology* (eds Hall, B. K. & Olson, W. M.) 35–53 (Harvard Univ. Press, 2003).
2. Kepecs, A. & Fishell, G. Interneuron cell types are fit to function. *Nature* **505**, 318–326 (2014).
3. Stegle, O., Teichmann, S. A. & Marioni, J. C. Computational and analytical challenges in single-cell transcriptomics. *Nat. Rev. Genet.* **16**, 133–145 (2015).
4. Shapiro, E., Biezuner, T. & Linnarsson, S. Single-cell sequencing-based technologies will revolutionize whole-organism science. *Nat. Rev. Genet.* **14**, 618–630 (2013).
5. Schwartzman, O. & Tanay, A. Single-cell epigenomics: techniques and emerging applications. *Nat. Rev. Genet.* **16**, 716–726 (2015).
6. Alberts, B. The cell as a collection of protein machines: preparing the next generation of molecular biologists. *Cell* **92**, 291–294 (1998).
7. Hartwell, L. H., Hopfield, J. J., Leibler, S. & Murray, A. W. From molecular to modular cell biology. *Nature* **402**, C47–C52 (1999).
8. Pereira-Leal, J. B., Levy, E. D. & Teichmann, S. A. The origins and evolution of functional modules: lessons from protein complexes. *Phil. Trans. R. Soc. B* **361**, 507–517 (2006).
9. Pereira-Leal, J. B., Levy, E. D., Kamp, C. & Teichmann, S. A. Evolution of protein complexes by duplication of homomeric interactions. *Genome Biol.* **8**, R51 (2007).
10. Achim, K. & Arendt, D. Structural evolution of cell types by step-wise assembly of cellular modules. *Curr. Opin. Genet. Dev.* **27**, 102–108 (2014).
11. Vickaryous, M. K. & Hall, B. K. Human cell type diversity, evolution, development, and classification with special reference to cells derived from the neural crest. *Biol. Rev.* **81**, 425–455 (2006).
The first account of human cell type diversity constructed through the lens of evolution.
12. Musser, J. M. & Wagner, G. P. Character trees from transcriptome data: origin and individuation of morphological characters and the so-called 'species signal'. *J. Exp. Zool.* **324**, 588–604 (2015).
Introduces the phenomenon of concerted transcriptome evolution and shows how it influences the comparison of cell type transcriptomes across species.
13. Valentine, J. W., Collins, A. G. & Meyer, C. P. Morphological complexity increase in metazoans. *Paleobiology* **20**, 131–142 (1994).
14. Wagner, G. P., Kin, K., Muglia, L. & Pavli, M. Evolution of mammalian pregnancy and the origin of the decidual stromal cell. *Int. J. Dev. Biol.* **58**, 117–126 (2014).
15. Schlosser, G. Modularity and the units of evolution. *Theory Biosci.* **121**, 1–80 (2002).
16. Wagner, G. P. & Lynch, V. J. Evolutionary novelties. *Curr. Biol.* **20**, R48–R52 (2010).
17. Wagner, G. P. *Homology, Genes, And Evolutionary Innovation* (Princeton Univ. Press, 2014).
A modern account of the molecular underpinnings of homology and evolutionary novelty.
18. Pavlicev, M., Wagner, G. P., Noonan, J. P., Hallgrimsson, B. & Cheverud, J. M. Genomic correlates of relationship QTL involved in fore- versus hind limb divergence in mice. *Genome Biol. Evol.* **5**, 1926–1936 (2013).
19. Arendt, D. The evolution of cell types in animals: emerging principles from molecular studies. *Nat. Rev. Genet.* **9**, 868–882 (2008).
Introduces the sister cell type model for the origin of novel cell types and sets the stage for an evolutionary perspective on the cell type concept.
20. Serb, J. M. & Oakley, T. H. Hierarchical phylogenetics as a quantitative analytical framework for evolutionary developmental biology. *Bioessays* **27**, 1158–1166 (2005).
Shows how phylogenetic tools can be used to track the history and interrelationships of cell types and other organismal parts.
21. Arendt, D. Evolution of eyes and photoreceptor cell types. *Int. J. Dev. Biol.* **47**, 563–571 (2003).
22. Brunet, T. *et al.* The evolutionary origin of bilaterian smooth and striated myocytes. Preprint at bioRxiv http://dx.doi.org/10.1101/064881 (2016).
23. Graf, T. & Enver, T. Forcing cells to change lineages. *Nature* **462**, 587–594 (2009).
A pioneering paper that suggests that the gene regulatory network underlying cell type identity has a flat hierarchy with a few transcription factor genes controlling many downstream effector genes.
24. Hobert, O. Regulatory logic of neuronal diversity: terminal selector genes and selector motifs. *Proc. Natl Acad. Sci. USA* **105**, 20067–20071 (2008).
25. Saint-André, V. *et al.* Models of human core transcriptional regulatory circuitries. *Genome Res.* **26**, 385–396 (2016).
26. Young, R. A. Control of the embryonic stem cell state. *Cell* **144**, 940–954 (2011).
27. Flames, N. & Hobert, O. Gene regulatory logic of dopamine neuron differentiation. *Nature* **458**, 885–889 (2009).
28. Hobert, O. Regulation of terminal differentiation programs in the nervous system. *Annu. Rev. Cell Dev. Biol.* **27**, 681–696 (2011).
29. Oosterveen, T. *et al.* SoxB1-driven transcriptional network underlies neural-specific interpretation of morphogen signals. *Proc. Natl Acad. Sci. USA* **110**, 7330–7335 (2013).
30. Mullen, A. C. *et al.* Master transcription factors determine cell-type-specific responses to TGF-β signaling. *Cell* **147**, 565–576 (2011).
Demonstrates empirically that cell type terminal selectors are responsible for mediating the cell type-specific response to a common cellular signal.
31. Trompouki, E. *et al.* Lineage regulators direct BMP and Wnt pathways to cell-specific programs during differentiation and regeneration. *Cell* **147**, 577–589 (2011).
32. Erwin, D. & Davidson, E. The evolution of hierarchical gene regulatory networks. *Nat. Rev. Genet.* **10**, 141–148 (2009).
33. Deneris, E. S. & Hobert, O. Maintenance of postmitotic neuronal cell identity. *Nat. Neurosci.* **17**, 899–907 (2014).
34. Davis, R. L., Weintraub, H. & Lassar, A. B. Expression of a single transfected cDNA converts fibroblasts to myoblasts. *Cell* **51**, 987–1000 (1987).
35. Arlotta, P. & Hobert, O. Homeotic transformations of neuronal cell identities. *Trends Neurosci.* **38**, 751–762 (2015).
36. Hobert, O. Terminal selectors of neuronal identity. *Curr. Top. Dev. Biol.* **116**, 455–475 (2016).
Outlines the important role of terminal selector genes in establishing and maintaining cell type identity.
37. Lee, S. *et al.* A regulatory network to segregate the identity of neuronal subtypes. *Dev. Cell* **14**, 877–889 (2008).
38. Duggan, A., Ma, C. & Chalfie, M. Regulation of touch receptor differentiation by the *Caenorhabditis elegans mec-3* and *unc-86* genes. *Development* **125**, 4107–4119 (1998).
39. Rudra, D. *et al.* Transcription factor Foxp3 and its protein partners form a complex regulatory network. *Nat. Immunol.* **13**, 1010–1019 (2012).
40. Gordon, P. M. & Hobert, O. A competition mechanism for a homeotic neuron identity transformation in *C. elegans*. *Dev. Cell* **34**, 206–219 (2015).
41. Jin, Z. *et al.* Different transcription factors regulate nestin gene expression during P19 cell neural differentiation and central nervous system development. *J. Biol. Chem.* **284**, 8160–8173 (2009).
42. Lodato, M. A. *et al.* SOX2 co-occupies distal enhancer elements with distinct POU factors in ESCs and NPCs to specify cell state. *PLoS Genet.* **9**, e1003288 (2013).
43. Sokolik, C. *et al.* Transcription factor competition allows embryonic stem cells to distinguish authentic signals from noise. *Cell Syst.* **1**, 117–129 (2015).
44. 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).
45. Henke, R. M. *et al.* Neurog2 is a direct downstream target of the Ptf1a–Rbpj transcription complex in dorsal spinal cord. *Development* **136**, 2945–2954 (2009).
46. Borromeo, M. D. *et al.* A transcription factor network specifying inhibitory versus excitatory neurons in the dorsal spinal cord. *Development* **141**, 3102–3102 (2014).
47. He, M. *et al.* Cell-type-based analysis of microRNA profiles in the mouse brain. *Neuron* **73**, 35–48 (2012).
48. Hobert, O. Gene regulation by transcription factors and microRNAs. *Science* **319**, 1785–1786 (2008).
49. Jovičič, A. *et al.* Comprehensive expression analyses of neural cell-type-specific miRNAs identify new determinants of the specification and maintenance of neuronal phenotypes. *J. Neurosci.* **33**, 5127–5137 (2013).
50. Poole, R. J. & Hobert, O. Early embryonic programming of neuronal left/right asymmetry in *C. elegans*. *Curr. Biol.* **16**, 2279–2292 (2006).
51. Kent, O. A., McCall, M. N., Cornish, T. C. & Halushka, M. K. Lessons from miR-143/145: the importance of cell-type localization of miRNAs. *Nucleic Acids Res.* **42**, 7528–7538 (2014).
52. Irimia, M. *et al.* A highly conserved program of neuronal microexons is misregulated in autistic brains. *Cell* **159**, 1511–1523 (2014).
53. Charlet-B, N., Logan, P., Singh, G. & Cooper, T. A. Dynamic antagonism between ETR-3 and PTB regulates cell type-specific alternative splicing. *Mol. Cell* **9**, 649–658 (2002).
54. Feng, Y. *et al.* SRp38 regulates alternative splicing and is required for Ca^{2+} handling in the embryonic heart. *Dev. Cell* **16**, 528–538 (2009).
55. Chen, M. & Manley, J. L. Mechanisms of alternative splicing regulation: insights from molecular and genomics approaches. *Nat. Rev. Mol. Cell Biol.* **10**, 741–754 (2009).
56. Giampietro, C. *et al.* The alternative splicing factor Nova2 regulates vascular development and lumen formation. *Nat. Commun.* **6**, 8479 (2015).
57. Lauri, A. *et al.* Development of the annelid axochord: insights into notochord evolution. *Science* **345**, 1365–1368 (2014).
58. Lamb, T. D. Evolution of phototransduction, vertebrate photoreceptors and retina. *Prog. Retin. Eye Res.* **36**, 52–119 (2013).
59. Gellersen, B. & Brosens, J. J. Cyclic decidualization of the human endometrium in reproductive health and failure. *Endocr. Rev.* **35**, 851–905 (2014).
60. Emera, D., Romero, R. & Wagner, G. The evolution of menstruation: a new model for genetic assimilation. *Bioessays* **34**, 26–35 (2011).
61. Kin, K., Nnamani, M. C., Lynch, V. J., Michaelides, E. & Wagner, G. P. Cell-type phylogenetics and the origin of endometrial stromal cells. *Cell Rep.* **10**, 1398–1409 (2015).
62. Mess, A. & Carter, A. M. Evolutionary transformations of fetal membrane characters in Eutheria with special reference to Afrotheria. *J. Exp. Zool.* **306**, 140–163 (2006).
63. Chavan, A. R., Bhullar, B.-A. S. & Wagner, G. P. What was the ancestral function of decidual stromal cells? A model for the evolution of eutherian pregnancy. *Placenta* **40**, 40–51 (2016).
64. Lynch, V. J., Brayer, K., Gellersen, B. & Wagner, G. P. HoxA-11 and FOXO1A cooperate to regulate decidual prolactin expression: towards inferring the core transcriptional regulators of decidual genes. *PLoS ONE* **4**, e6845 (2009).

65. Kao, L. C. *et al.* Global gene profiling in human endometrium during the window of implantation. *Endocrinology* **143**, 2119–2138 (2002).
66. Eda Akbas, G., Song, J. & Taylor, H. S. A HOXA10 estrogen response element (ERE) is differentially regulated by 17 beta-estradiol and diethylstilbestrol (DES). *J. Mol. Biol.* **340**, 1013–1023 (2004).
67. Sarno, J. L., Kliman, H. J. & Taylor, H. S. HOXA10, Pbx2, and Meis1 protein expression in the human endometrium: formation of multimeric complexes on HOXA10 target genes. *J. Clin. Endocrinol. Metab.* **90**, 522–528 (2005).
68. Wetendorf, M. & DeMayo, F. J. The progesterone receptor regulates implantation, decidualization, and glandular development via a complex paracrine signaling network. *Mol. Cell. Endocrinol.* **357**, 108–118 (2012).
69. Lynch, V. J. *et al.* Adaptive changes in the transcription factor HoxA-11 are essential for the evolution of pregnancy in mammals. *Proc. Natl Acad. Sci. USA* **105**, 14928–14933 (2008).
Reconstructs ancestral transcription factors to show the importance of transcription factor protein-coding changes for cell type evolution.
70. Lynch, V. J., May, G. & Wagner, G. P. Regulatory evolution through divergence of a phosphoswitch in the transcription factor CEBPB. *Nature* **480**, 383–386 (2012).
71. Nnamani, M. C. *et al.* A derived allosteric switch underlies the evolution of conditional cooperativity between HOXA11 and FOXO1. *Cell Rep.* **15**, 2097–2108 (2016).
72. Brayer, K. J., Lynch, V. J. & Wagner, G. P. Evolution of physical interactions among the transcription factors HoxA-11 and FOXO1a during the evolution of pregnancy in mammals. *Soc. Integr. Comp. Biol.* **49**, E21 (2009).
73. Ono, H., Kozmik, Z., Yu, J.-K. & Wada, H. A novel N-terminal motif is responsible for the evolution of neural crest-specific gene-regulatory activity in vertebrate FoxD3. *Dev. Biol.* **385**, 396–404 (2014).
74. Nitta, K. R. *et al.* Conservation of transcription factor binding specificities across 600 million years of bilateria evolution. *eLife* **4**, e04837 (2015).
75. Kent, M. L. *et al.* Recent advances in our knowledge of the Myxozoa. *J. Eukaryot. Microbiol.* **48**, 395–413 (2001).
76. Liang, C., 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).
77. Owen, R. *On the Archetype and Homologies of the Vertebrate Skeleton* (John van Voorst, 1848).
78. Steinmetz, P. R. H. *et al.* Independent evolution of striated muscles in cnidarians and bilaterians. *Nature* **487**, 231–234 (2012).
79. Stefanakis, N., Carrera, I. & Hobert, O. Regulatory logic of pan-neuronal gene expression in *C. elegans*. *Neuron* **87**, 733–750 (2015).
A systematic investigation of how different neuronal cell types use both cell type-specific and more general regulatory information to express pan-neuronal genes.
80. Shalchian-Tabrizi, K. *et al.* Multigene phylogeny of Choanozoa and the origin of animals. *PLoS ONE* **3**, e2098 (2008).
81. Philippe, H. *et al.* Phylogenomics revives traditional views on deep animal relationships. *Curr. Biol.* **19**, 706–712 (2009).
82. Dunn, C. *et al.* Broad phylogenomic sampling improves resolution of the animal tree of life. *Nature* **452**, 745–749 (2008).
83. Nickel, M. Evolutionary emergence of synaptic nervous systems: what can we learn from the non-synaptic, nerveless Porifera? *Invertebr. Biol.* **129**, 1–16 (2010).
84. Jones, W. C. Is there a nervous system in sponges? *Biol. Rev.* **37**, 1–50 (1962).
85. Simpson, T. L. *The Cell Biology of Sponges* (Springer Verlag, 1984).
86. Mackie, G. O. The elementary nervous system revisited. *Am. Zool.* **30**, 907–920 (1990).
87. Meech, R. W. & Mackie, G. O. in *Invertebrate Neurobiology* (eds North, G. & Greenspann, J.) 581–615 (Cold Spring Harbor Laboratory Press, 2007).
88. Leys, S. P. in *Porifera Research: Biodiversity, Innovation and Sustainability* (eds Custódio, M., Lôbo-Hajdu, G., Hajdu, E. & Muricy, G.) 53–59 (Museu Nacional, 2007).
89. Smith, C. L. *et al.* Novel cell types, neurosecretory cells, and body plan of the early-diverging metazoan *Trichoplax adhaerens*. *Curr. Biol.* **24**, 1565–1572 (2014).
90. Alié, A. & Manuel, M. The backbone of the post-synaptic density originated in a unicellular ancestor of choanoflagellates and metazoans. *BMC Evol. Biol.* **10**, 34 (2010).
91. Sakarya, O. *et al.* A post-synaptic scaffold at the origin of the animal kingdom. *PLoS ONE* **2**, e506 (2007).
92. Kienle, N., Kloepper, T. H. & Fasshauer, D. Differences in the SNARE evolution of fungi and metazoa. *Biochem. Soc. Trans.* **37**, 787–791 (2009).
93. Burkhardt, P. *et al.* Primordial neurosecretory apparatus identified in the choanoflagellate *Monosiga brevicollis*. *Proc. Natl Acad. Sci. USA* **108**, 15264–15269 (2011).
94. Leys, S. P. & Riesgo, A. Epithelia, an evolutionary novelty of metazoans. *J. Exp. Zool. Mol. Dev. Evol.* **318**, 438–447 (2012).
95. Grimson, M. J. *et al.* Adherens junctions and β-catenin-mediated cell signalling in a non-metazoan organism. *Nature* **408**, 727–731 (2000).
96. Murray, P. S. & Zaidel-Bar, R. Pre-metazoan origins and evolution of the cadherin adhesome. *Biol. Open* **3**, 1183–1195 (2014).
97. Miller, P. W., Clarke, D. N., Weis, W. I., Lowe, C. J. & Nelson, W. J. The evolutionary origin of epithelial cell–cell adhesion mechanisms. *Curr. Top. Membr.* **72**, 267–311 (2013).
98. Bamji, S. X. *et al.* Role of beta-catenin in synaptic vesicle localization and presynaptic assembly. *Neuron* **40**, 719–731 (2003).
99. Jungling, K. *et al.* N-Cadherin transsynaptically regulates short-term plasticity at glutamatergic synapses in embryonic stem cell-derived neurons. *J. Neurosci.* **26**, 6968–6978 (2006).
100. Togashi, H. *et al.* Cadherin regulates dendritic spine morphogenesis. *Neuron* **35**, 77–89 (2002).
101. Elia, L. P., Yamamoto, M., Zang, K. & Reichardt, L. F. p120 catenin regulates dendritic spine and synapse development through Rho-family GTPases and cadherins. *Neuron* **51**, 43–56 (2006).
102. Bozdagi, O. *et al.* Persistence of coordinated long-term potentiation and dendritic spine enlargement at mature hippocampal CA1 synapses requires N-cadherin. *J. Neurosci.* **30**, 9984–9989 (2010).
103. Mendez, P., De Roo, M., Poglia, L., Klauser, P. & Muller, D. N-Cadherin mediates plasticity-induced long-term spine stabilization. *J. Cell Biol.* **189**, 589–600 (2010).
104. Conaco, C. *et al.* Functionalization of a protosynaptic gene expression network. *Proc. Natl Acad. Sci. USA* **109** (Suppl. 1), 10612–10618 (2012).
105. Ori, A. *et al.* Spatiotemporal variation of mammalian protein complex stoichiometries. *Genome Biol.* **17**, 47 (2016).
Provides an expansive look into cell type-specific variation of human protein complexes, suggesting most complexes vary in their composition.
106. Weimer, C. *et al.* Differential roles of ArfGAP1, ArfGAP2, and ArfGAP3 in COPI trafficking. *J. Cell Biol.* **183**, 725–735 (2008).
107. Kartberg, F. *et al.* ARFGAP2 and ARFGAP3 are essential for COPI coat assembly on the Golgi membrane of living cells. *J. Biol. Chem.* **285**, 36709–36720 (2010).
108. Emes, R. D. *et al.* Evolutionary expansion and anatomical specialization of synapse proteome complexity. *Nat. Neurosci.* **11**, 799–806 (2008).
109. Ryan, T. J. & Grant, S. G. N. The origin and evolution of synapses. *Nat. Rev. Neurosci.* **10**, 701–712 (2009).
110. Emes, R. D. & Grant, S. G. N. Evolution of synapse complexity and diversity. *Annu. Rev. Neurosci.* **35**, 111–131 (2012).
111. Nithianantharajah, J. *et al.* Synaptic scaffold evolution generated components of vertebrate cognitive complexity. *Nat. Neurosci.* **16**, 1–11 (2012).
112. Kin, K. *et al.* The transcriptomic evolution of mammalian pregnancy: gene expression innovations in endometrial stromal fibroblasts. *Genome Biol. Evol.* **8**, 2459–2473 (2016).
113. Parsons, M. & Adams, J. C. Rac regulates the interaction of fascin with protein kinase C in cell migration. *J. Cell Sci.* **121**, 2805–2813 (2008).
114. Lin-Jones, J. & Burnside, B. Retina-specific protein fascin 2 is an actin cross-linker associated with actin bundles in photoreceptor inner segments and calycal processes. *Invest. Ophthalmol. Vis. Sci.* **48**, 1380–1388 (2007).
115. Ohno, S. *Evolution by Gene Duplication* (Springer, 1970).
116. Lagman, D., Sundström, G., Daza, D. O., Abalo, X. M. & Larhammar, D. Expansion of transducin subunit gene families in early vertebrate tetraploidizations. *Genomics* **100**, 203–211 (2012).
117. Lagman, D. *et al.* The vertebrate ancestral repertoire of visual opsins, transducin alpha subunits and oxytocin/vasopressin receptors was established by duplication of their shared genomic region in the two rounds of early vertebrate genome duplications. *BMC Evol. Biol.* **13**, 238 (2013).
118. Fu, Y. & Yau, K.W. Phototransduction in mouse rods and cones. *Pflugers Arch.* **454**, 805–819 (2007).
119. Chen, C. K. *et al.* Replacing the rod with the cone transducin α subunit decreases sensitivity and accelerates response decay. *J. Physiol. (Lond.)* **588**, 3231–3241 (2010).
120. Kawamura, S. & Tachibanaki, S. Rod and cone photoreceptors: molecular basis of the difference in their physiology. *Comp. Biochem. Physiol. Physiol.* **150**, 369–377 (2008).
121. Tachibanaki, S., Yonetsu, S. I., Fukaya, S., Koshitani, Y. & Kawamura, S. Low activation and fast inactivation of transducin in carp cones. *J. Biol. Chem.* **287**, 41186–41194 (2012).
122. Korenbrot, J. I. Speed, sensitivity, and stability of the light response in rod and cone photoreceptors: facts and models. *Prog. Retin. Eye Res.* **31**, 442–466 (2012).
123. Reyer, R. W. Regeneration of the lens in the amphibian eye. *Q. Rev. Biol.* **29**, 1–46 (1954).
124. Sulston, J. E., Schierenberg, E., White, J. G. & Thomson, J. N. The embryonic cell lineage of the nematode *Caenorhabditis elegans*. *Dev. Biol.* **100**, 64–119 (1983).
125. Emerson, M. M., Surzenko, N., Goetz, J. J., Trimarchi, J. & Cepko, C. L. Otx2 and Onecut1 promote the fates of cone photoreceptors and horizontal cells and repress rod photoreceptors. *Dev. Cell* **26**, 59–72 (2013).
126. Bayraktar, O. A. & Doe, C. Q. Combinatorial temporal patterning in progenitors expands neural diversity. *Nature* **498**, 449–455 (2013).
127. Sternberg, P. W. & Horvitz, H. R. The genetic control of cell lineage during nematode development. *Annu. Rev. Genet.* **18**, 489–524 (1984).
128. Fritzsch, B., Beisel, K. W., Pauley, S. & Soukup, G. Molecular evolution of the vertebrate mechanosensory cell and ear. *Int. J. Dev. Biol.* **51**, 663–678 (2007).
129. Fritzsch, B. & Piatigorsky, J. Ancestry of photic and mechanic sensation? *Science* **308**,1113–1114 (2005).
130. Fritzsch, B., Beisel, K. W. & Bermingham, N. A. Developmental evolutionary biology of the vertebrate ear: conserving mechanoelectric transduction and developmental pathways in diverging morphologies. *Neuroreport* **11**, R35–R44 (2000).
131. Wang, Z., Young, R. L., Xue, H. & Wagner, G. P. Transcriptomic analysis of avian digits reveals conserved and derived digit identities in birds. *Nature* **477**, 583–586 (2011).
132. Nair, N. U. *et al.* A maximum-likelihood approach for building cell-type trees by lifting. *BMC Genomics* **17** (Suppl. 1), 14 (2016).
133. Pavlicev, M. & Widder, S. Wiring for independence: positive feedback motifs facilitate individuation of traits in development and evolution. *J. Exp. Zool.* **324**, 104–113 (2015).
134. Dueck, H. *et al.* Deep sequencing reveals cell-type-specific patterns of single-cell transcriptome variation. *Genome Biol.* **16**, 122 (2015).
135. Dueck, H., Eberwine, J. & Kim, J. Variation is function: are single cell differences functionally important? Testing the hypothesis that single cell variation is required for aggregate function. *Bioessays* **38**, 172–180 (2016).
136. Brunskill, E. W. *et al.* Single cell dissection of early kidney development: multilineage priming. *Development* **141**, 3093–3101 (2014).
137. McKenna, A. *et al.* Whole organism lineage tracing by combinatorial and cumulative genome editing. *Science* **353**, aaf7907 (2016).
138. Wang, S., Sengel, C., Emerson, M. M. & Cepko, C. L. A gene regulatory network controls the binary fate decision of rod and bipolar cells in the vertebrate retina. *Dev. Cell* **30**, 513–527 (2014).
139. Corbo, J. C., Myers, C. A., Lawrence, K. A., Jadhav, A. P. & Cepko, C. L. A typology of photoreceptor gene expression patterns in the mouse. *Proc. Natl Acad. Sci. USA* **104**, 12069–12074 (2007).
140. Kim, D. S. *et al.* Identification of molecular markers of bipolar cells in the murine retina. *J. Comp. Neurol.* **507**, 1795–1810 (2008).

141. 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).
142. Cepko, C. Intrinsically different retinal progenitor cells produce specific types of progeny. *Nat. Rev. Neurosci.* **15**, 615–627 (2014).
143. Centanin, L. *et al.* Exclusive multipotency and preferential asymmetric divisions in post-embryonic neural stem cells of the fish retina. *Development* **141**, 3472–3482 (2014).
144. Fujiyama, T. *et al.* Inhibitory and excitatory subtypes of cochlear nucleus neurons are defined by distinct bHLH transcription factors, Ptf1a and Atoh1. *Development* **136**, 2049–2058 (2009).
145. Yamada, M. *et al.* Specification of spatial identities of cerebellar neuron progenitors by ptf1a and atoh1 for proper production of GABAergic and glutamatergic neurons. *J. Neurosci.* **34**, 4786–4800 (2014).
146. Tosches, M. A. & Arendt, D. The bilaterian forebrain: an evolutionary chimaera. *Curr. Opin. Neurobiol.* **23**, 1080–1089 (2013).
147. Xiang, M. *et al.* Essential role of POU-domain factor Brn-3c in auditory and vestibular hair cell development. *Proc. Natl Acad. Sci. USA* **94**, 9445–9450 (1997).
148. Bermingham, N. A. *et al. Math1*: an essential gene for the generation of inner ear hair cells. *Science* **284**, 1837–1841 (1999).
149. Wallis, D. *et al.* The zinc finger transcription factor Gfi1, implicated in lymphomagenesis, is required for inner ear hair cell differentiation and survival. *Development* **130**, 221–232 (2003).
150. Eijkelkamp, N., Quick, K. & Wood, J. N. Transient receptor potential channels and mechanosensation. *Annu. Rev. Neurosci.* **36**, 519–546 (2013).
151. Schlosser, G. Vertebrate cranial placodes as evolutionary innovations — the ancestor's tale. *Curr. Top. Dev. Biol.* **111**, 235–300 (2015).
Reconstructs the evolutionary history of neurogenic placodes in vertebrates, and based on evolutionary and developmental evidence suggests a hierarchy of relatedness among different placode-derived organs.
152. Arendt, D., Benito-Gutierrez, E., Brunet, T. & Marlow, H. Gastric pouches and the mucociliary sole: setting the stage for nervous system evolution. *Philos. Trans. R. Soc. B* **370**, 20150286 (2015).
153. Satoh, T. & Fekete, D. M. Clonal analysis of the relationships between mechanosensory cells and the neurons that innervate them in the chicken ear. *Development* **132**, 1687–1697 (2005).
154. Simionato, E. *et al. atonal*- and *achaete-scute*-related genes in the annelid *Platynereis dumerilii*: insights into the evolution of neural basic-helix–loop–helix genes. *BMC Evol. Biol.* **8**, 170 (2008).
155. Baker, C. V. H., Modrell, M. S. & Gillis, J. A. The evolution and development of vertebrate lateral line electroreceptors. *J. Exp. Biol.* **216**, 2515–2522 (2013).
156. Whitear, M. Merkel cells in lower vertebrates. *Arch. Histol. Cytol.* **52** (Suppl.), 415–422 (1989).
157. Woo, S.-H., Lumpkin, E. A. & Patapoutian, A. Merkel cells and neurons keep in touch. *Trends Cell Biol.* **25**, 74–81 (2015).
158. Masuda, M. *et al.* Regulation of *POU4F3* gene expression in hair cells by 5' DNA in mice. *Neuroscience* **197**, 48–64 (2011).
159. Tang, J. W., Chen, J. S. & Zeller, R. W. Transcriptional regulation of the peripheral nervous system in *Ciona intestinalis*. *Dev. Biol.* **378**, 183–193 (2013).
160. Stolfi, A., Ryan, K., Meinertzhagen, I. A. & Christiaen, L. Migratory neuronal progenitors arise from the neural plate borders in tunicates. *Nature* **527**, 371–374 (2015).
161. Eisenhoffer, G. T. & Rosenblatt, J. Bringing balance by force: live cell extrusion controls epithelial cell numbers. *Trends Cell Biol.* **23**, 185–192 (2013).
162. Cowdry, E. V. *General Cytology: A Textbook of Cellular Structure and Function for Students of Biology and Medicine* (The Univ. of Chicago Press, 1924).
163. Wilson, E. B. *The Cell in Development and Heredity* (The Macmillan Company, 1925).
164. Harris, H. *The Cells of the Body: A History of Somatic Cell Genetics* (Cold Spring Harbor Laboratory Press, 1995).
165. Harris, H. *The Birth of the Cell* (Yale Univ. Press, 1999).
166. Hertwig, O. *Handbuch der Vergleichenden und Experimentellen Entwickelungslehre der Wirbeltiere* (in German) (Fischer, 1906).
167. Strasburger, E. & Hertwig, O. *Zellen- und Gewebelehre, Morphologie, und Entwicklungsgeschichte* (in German) (B. G. Teubner, 1913).
168. Maienschein, J. *Transforming Traditions in American Biology, 1880–1915* (Johns Hopkins Univ. Press, 1991).
169. Morgan, T. H. *The Mechanism of Mendelian Heredity* (Holt, 1915).
170. Boveri, T. *Die Organismen als Historische Wesen* (in German) (Kgl. Universitätsdruckerei von H. Stürtz, 1906).
An early conceptual account that introduces the idea that developmental differentiation and evolutionary differentiation must be a consequence of regulatory structures within the hereditary material.

Acknowledgements
This paper resulted from discussions of a working group at the Santa Fe Institute (SFI), New Mexico, USA, sponsored by the Arizona State University (ASU)-SFI Center for Biosocial Complex Systems. The authors are grateful for their support. Research on neuron type evolution in the Arendt laboratory is financially supported by a European Research Council Advanced grant (294810 Brain Evo-Devo). Research on cell type origination in the Wagner laboratory is financially supported by a grant from the John Templeton Foundation (grant no. 54860). The opinions expressed in this article are not those of the John Templeton Foundation.

Competing interests statement
The authors declare no competing interests.