Structural evolution of cell types by step-wise assembly of cellular modules
Kaia Achim and Detlev Arendt

Cell types are composed of cellular modules exerting specific subfunctions. The evolutionary emergence and diversification of these modules can be tracked through the comparative analysis of genomes. Here, we survey recent advances elucidating the origin of neurons, smooth and striated muscle cells and of the T- and B-cells of the immune system in the diverging lineages of animal evolution. Gene presence and absence analyses in various metazoan genomes allow mapping the step-wise assembly of key modules – such as the postsynaptic density characteristic for neurons or the z-disk characteristic for striated muscle – on the animal evolutionary tree. Using this approach, first insight into the structural evolution of cell types can be gained.

Addresses
European Molecular Biology Laboratory, Developmental Biology Unit, Meyerhofstrasse 1, 69012 Heidelberg, Germany

Corresponding author: Arendt, Detlev (arendt@embl.de)

Introduction
Animals are composed of cell types of distinct structure and function. Epithelial cell types provide barriers between environments; muscle cell types contain contractile filaments enabling all sorts of movement; neuron types with their dendrites and axons allow directed information transfer via synapses; sensory cell types read environmental cues; and immune cells with their multitude of specific and unspecific receptors constitute the organismal defence system.

What is a cell type? In essence, the definition of a cell type is structural. It refers to a specific phenotype, or ‘morphotype’ [1], of differentiated cells in the organismal context. Obviously, the cell type structure is a manifestation of its molecular composition, adapted to specific functions. Typically, cellular functions require the cooperation of many proteins and other biomolecules that constitute ‘modules’ [2–4]. We can thus envisage a cell type as an assembly of modules exerting discrete subfunctions. For example, a sensory or motile cilium, or the actomyosin contractile machinery is a cellular module; an assembly of membrane channels that enables action potentials is a module, as are the various signalling cascades. Modularity is clearly favoured in evolution, as it facilitates the adaptive variation of one module without perturbing the other and thus increases fitness in changing environments [5,6].

How did the diversity of cell types and modules evolve? All cell types that constitute today’s multicellular organisms result from the step-wise diversification of their unicellular founders. Starting with one, then few cell types, these diversified more and more in the various evolutionary lineages. Cell types that evolved from the same immediate precursor in a given lineage are referred to as sister cell types [7]. If, in two emergent sister cell types, the cellular modules are basically retained (but modified to some extent), structure and function of these cell types initially remain the same but diverge with time. Good examples for such ‘divergence of function’ are the rods and cones of the vertebrate retina (that both retained and modified the ciliary photoreceptor module in different directions [7]). If, instead, cellular modules are lost in one or both sister cell types, structure and function of these cell types become distinct. Illustrating such ‘segregation of function’, the bipolar cells of the vertebrate retina appear to have lost the photoreceptor module that was present in their ancient precursors [7,8]. This process is also referred to as ‘division of labour’ [9**]. Notably, divergence and segregation of function can co-occur in the same diversification event [7].

Can we track the evolutionary process that gave rise to today’s diversity of modules and cell types? Given that the assembly of all cellular modules follows information encoded in the genome, comparative genomics has great potential in unravelling the genealogies of these modules. In particular, genome sequencing allows inferring when a given module has come into place in the course of evolution; we can then infer, from the cell type(s) present in these ancestors, what the first function of this module has been and how this relates to the later functions exerted by the module; furthermore, sequence comparisons reveal whether similar modules in distinct evolutionary lineages are the result of homology or convergence.
The evolution of cell types is written in the genomes. Achim and Arendt

Figure 1

Evolution of modules and emergence of cell types in nervous system (upper panel) and musculature (lower panel). Blue boxes indicate the gain of a subcellular module or its constituting proteins; black boxes indicate the emergence of a new cell type. Boxes connected by a dashed black line indicate convergent evolution of similar traits. Animal drawings were taken from Nielsen [57] and other sources [URL: http://www.examguideonline.com (sea anemone), URL: http://www.uni-giessen.de/~gf1307/ (Platynereis dumerilii) drawing by Ursula Fischer].
Our minireview surveys recent comparative genomics studies that address the evolution of key modules of neurons [10,11,12*,13] and muscle cells [14**], such as synapses and acto-myosin filaments, and of cell types of the immune system [15**,16**]. These studies track the emergence of the molecular components that constitute cell type specific modules through animal evolution and provide excellent case studies for functional divergence and division of labour. Furthermore, they exemplify a general principle that appears to govern cell type evolution: that, in many cases, novel cell types such as neurons and myocytes evolve by specialized usage of pre-existing modules rather than by the de novo-emergence of new modules. Illustrating this, our Figure 1 maps the gradual emergence of key cellular modules antedating the emergence of neurons and muscle cells on a simplified animal evolutionary tree, as deduced from these studies.

**Evolution of neurons**

As the greatest diversity of cell types is found in the nervous system [17], understanding the evolution of neural cell types is key to understanding animal complexity. The recent annotation of basal metazoan genomes [11,18–20] has revealed part lists of important neural modules that allow step-wise tracking of their evolutionary emergence. In this exercise, the modules of the chemical synapse are of particular interest as they allow tracking the origin of bona fide neurons, defined by their capacity to signal to individual target cells via synapses (Figure 1a).

Surprisingly, multiple genes encoding proteins of the highly complex postsynaptic density have recently been traced back to the choanoflagellate-metazoan ancestor [10]. As synapses are obviously absent in choanoflagellates (and in sponges and placozoans), these data indicate that, in early metazoa, this module must have served another function, before it became part of the synapse. Intriguingly, other studies suggest that the postsynaptic module indeed first acted as a ‘chemosensory module’ [21–24]: Initially sensing environmental cues (such as the amino acid glutamate indicating prey) the partaking receptors and ion channels may have started to receive internal information (such as the transmitter glutamate) from within the newly evolving synapse. Figure 2 illustrates how the postsynapse might have evolved from the chemosensory module [24]. In this scenario, the resulting sensory cell and neuron represent sister cell types; the different usage of chemosensory apparatus and postsynapse represents a divergence of function; and the specialization on sensory versus integrative functions is a division of labour event. Corroborating this scenario, ionotropic glutamate receptor families existed before the divergence of animals and plants and metabotropic glutamate (and GABA) receptors predate the metazoan radiation [11,12*] (Figure 1a); and, notably, both families are known to comprise chemosensors for external glutamate [25–27].

If, as these studies suggest, the postsynaptic module evolved from an ancient chemosensory module, when did this happen? The key step here seems to be the emergence of Neuroligin (Nlgn), the ligand mediating
the ‘handshake’ between pre- and postsynaptic neurons on the post-synaptic side. Nlgn has not been found in basal metazoans that lack neurons such as sponges [18,28] and the placozoaan Trichoplax [10,11], while it is present in the sea anemone Nematoctella that possesses neurons [10,28]. However, to illustrate a caveat of presence/absence analyses, Nlgn has not been found in the freshwater polyp Hydra, which possesses neurons [10]. As Hydra belongs to the cnidarians, this absence is necessarily due to secondary loss or strong modification (or the gene simply has not been found yet). The same might be true for the comb jelly Mnemiopsis that likewise possesses neurons with highly characteristic synapses [29] but apparently misses Nlgn. Interestingly, the presynaptic binding partner of Nlgn, neurexin, has been detected in Mnemiopsis [11]. Together, these data suggest that synapses evolved once and existed in the last common ancestor of cnenophores, cnidarians and bilaterians (Figure 1a), which would imply homology of neurons. To track the assembly of synapses further, it will be rewarding to similarly follow the emergence of proteins known to structurally assemble the presynaptic active zone and regulate synaptic vesicle release, such as the PDZ domain proteins ERC and RIM that are missing in sponges [18] but conserved across bilaterians [13].

Regarding the evolution of neurotransmitter systems, a genomic inventory of receptors, channels and synthesizing enzymes in sea anemone has revealed that acetylcholine, GABA/glycine, neuropeptide and hormone signalling likewise predates the last common ancestor of cnidarians and bilaterians [30]. Complementing this, a recent clustering analysis of neuropeptides and G-protein-coupled neuropeptide receptors shows that the emergence of the neuropeptide/GPCR signalling system predates the divergence of placozoans and identified a minimum of five neuropeptide/hormone signalling systems that were active in cnidarian-bilaterian ancestor [31*,32,33] (Figure 1a). Finally, pan-neuronal genes encoding RNA-binding proteins elav and Musashi, are present in sponges [18]. Binding to intronic sequences and 3′UTR sequences, elav proteins regulate alternative splicing and mRNA levels of neural genes [34]; interestingly, different human elav paralogs have recently been shown to regulate components of the glutamate synthesis pathway [34].

Evolution of smooth and striated muscle cell types

The various kinds of specialized muscle cell types in bilaterians are assumed to have evolved from contractile epithelial muscle cells [7,35]. Cells relating to such hypothetical muscle cell precursors, so-called myoepithelial cells, exist in extant cnidarians [36,37]. These cells have long basal contractile processes that resemble muscle fibres [37]. On the basis of electron optics, smooth and striated muscle cell types can be distinguished; both types are present in bilaterians and cnidarians [37,38] (Figure 1b). In cnenophores, most muscle cells lack the striation pattern (with the exception of the tentacle muscles in one species, Euplokamis) [39,40].

In an attempt to elucidate the evolution and interrelationship of smooth and striated muscle cell types in metazoans, Steinmetz and co-authors have recently mined genomic information from several early branching metazoans [14**]. They first establish that the core contractile module, the acto-myosin filament (comprising actin, myosin, tropomyosin and calmodulin), predates the metazoan radiation [14**] (Figure 1b); the first function of this module was in basic cell biological processes involving cytoskeletal remodeling [41]. Likewise, two duplicates of the myosin heavy chain that co-existed in unicellular ancestors, most likely conveyed different speeds of contraction [14**]. In early metazoans, myosin light chain kinase was added to the acto-myosin module that coupled the regulation of contractions to the intracellular concentration of calcium ions [14**]. All these events antedate the birth of smooth muscle cells that most likely occurred once (Figure 1b). Interestingly, the same study reveals that another cellular module specific for striated muscle cells, the z-disc, appears to have been recruited independently in bilaterians, cnidarians and cnenophores (Figure 1b, dashed line), as revealed by the absence of most bilaterian z-disc proteins in cnidarians [14**]. Notably, the striated muscle cells independently recruited the same ‘fast’ myosin heavy chain molecule for efficient contraction [14**].

Cell type diversification in the immune system

The vertebrate adaptive immune system provides another interesting case study for cell type evolution. It comprises two highly specialized cell types, the B and T lymphocytes. Upon antigen presentation, activated T lymphocytes can differentiate into cytotoxic (Tc) or helper (Th) cells. The latter amplify the response of B and Tc cells but also that of the macrophages, thus linking the adaptive and innate immune response. In addition, vertebrates also possess atypical, gamma/delta receptor-expressing T cells that can carry out various functions at the interface of adaptive and innate immunity. To elucidate the origin of protein modules characteristic for the adaptive immune response, recent studies analysed genomic information of basal vertebrates.

Curiously, lymphocytes in basal versus more advanced vertebrate lineages express different T-cell receptor coreceptors for target recognition: immunoglobulin (Ig)-repeat-containing CD receptors versus leucine-rich repeat containing variable lymphocyte receptors (VLR), respectively [42]. At first sight, this might indicate convergent evolution of T-cell lineages in these groups; however, a recent comparison of regulatory signatures reveals that, despite these differences, gnathostome and
cyclostome differentiating lymphocytes express similar cell type-specific combinations of transcription factors and membrane markers [16**]. These data suggest that two types of T-cells (Tc/Th cell-like, and ‘atypical’ T cell) and one type of B-cells already existed in the last common ancestor of all vertebrates.

Genome mining in the elephant shark and some other cartilaginous fishes has provided further clues on the diversification of T cell lineages. Namely, all components required for Tc cell development, but not those characteristic for the Th cell, were found in this basal vertebrate lineage [15**]. This would suggest that different modules enabling different modes of immunity were acquired by T lymphocytes at different times of evolution [15**]. The last T-cell types to appear in adaptive immunity evolution were probably the regulatory and memory T-cell [42,43] that emerged in bony vertebrate lineage, after the expansion of T cell activating interleukin family together with T-cell surface receptors and the emergence of key transcription factors in T cell subtype specification, such as RORC or FoxP3 [15**,44].

What about early stages of immune cell type evolution? In general, among the invertebrate coelomocytes (cells floating in the coelom), granular hemocytes (granulocytes) are considered homologous to vertebrate adaptive immune cells [7,45]. Invertebrate blood cells have been subclassified by morphological criteria, but are widely viewed as stage- or organismal state-specific descendants of the same lineage [45]. In the light of the notion of three immune cell types at the base of the vertebrate lineage, it will be interesting to assess when this divergence occurred. The availability of extensive molecular and morphological fingerprint catalogues of human and mouse blood cell types [46,47] will enable high-resolution comparisons with any cell-type specific transcriptomic data on the invertebrate side.

**Future prospects**

The identification of cellular modules in the various animal genomes and the mapping of components constituting these modules on the animal tree, as exemplified for the vertebrate stem line in Figure 1, provide an exciting new view of phenotypic evolution. With time, a comprehensive view on the modules present at specific nodes of the tree will emerge. In a pioneer study, Wenger and Galliot have recently identified four ‘hot spots’ of protein innovation on the evolutionary lineage leading to the vertebrates [48**]. Once the identified structural proteins that evolved during these innovation periods are fully understood and sorted into modules, this will result in a refined picture of the complexity of the respective ancestors.

Yet, the power of comparative genomics in reconstructing the evolution of cellular modules and cell types necessarily faces its limits. In many cases, the mere presence of a protein in a given genome will not be sufficient to assign it to a specific cellular module (unless biochemical or other relevant data is already available). Also, in many cases the presence of a module will also not suffice to attribute it to the diverse cell type(s) present in each animal. In most studies discussed here, this link has been (tentatively) established by wholemount in situ expression analysis of selected genes; for example, co-expression of the postsynaptic density module with the ‘neurogenic’ genes in the sponge Amphimedon reveals its presence in sensory cells [28,49]; or, although the genes for vertebrate Z-disk proteins alpha-actinin, muscleLIM and Ldhβ are present in cnidarians, they are not co-expressed in the striated muscle cells [14**], which indicates that the latter evolved convergently (see above). However, in some species hybridisation protocols are not available; and simultaneous co-labeling of animals with probes detecting transcripts of two or more genes is tedious and will be impossible in many cases.

In this context, single cell transcriptomics provides an exciting new opportunity for unbiased and quantitative characterization of cell types [50]. Single cell analysis techniques are rapidly improving, for example by combination of microfluidic platforms and advanced next generation sequencing techniques [51,52]. Expanding these protocols to representatives of the evolutionary lineages depicted in our Figure 1 will be especially rewarding for reconstructing cell type evolution of basal metazoans. Single cell transcriptomics will also contribute to unravel the specific combinations of transcription factors acting upstream of the cellular modules. A growing body of evidence indicates that genes encoding protein modules are often co-regulated by limited number of transcription factors (‘selector genes’), such as LIM and POU homeodomain family proteins [53**,54]; these factors act via similar cis-regulatory elements, thus forming so-called ‘programming modules’ [55,56].

Once sets of genes encoding cellular modules and their specifying transcription factors will be attributed, at larger scale, to specific cell types in different species, this will set the stage for the identification of homologous cell types. Also, it will be possible to elucidate sister cell type relationships within a given species. We predict that the combination of comparative genomics and comparative single cell-transcriptomics will boost our understanding of cell type evolution in animals.

**References and recommended reading**

Papers of particular interest, published within the period of review, have been highlighted as:

- of special interest
- of outstanding interest

The evolution of cell types is written in the genomes Achim and Arendt 107


14. Steinmetz PR, Kraus JE, Larroux C, Hammel JU, Amon- 


52. Tan SJ, Phan H, Gerry BM, Kuhn A, Hong LZ, Min Ong Y, Poon PS, Unger MA, Jones RC, Quake SR et al.: A microfluidic device for preparing next generation DNA sequencing libraries and for automating other laboratory protocols that require one or more column chromatography steps. PLoS One 2013, 8:e64084.


A milestone study showing that the glutamatergic neuron marker, VGLUT, is regulated by distinct cis-regulatory modules in different types of glutamatergic neurons in the nematode C. elegans. This reveals that terminal selector genes specify different cell type identities rather than specific cellular modules; and that a given module can be controlled by different selector genes depending on the cell type.


