Published in final edited form as:

Trends Neurosci. 2010 October ; 33(10): 435–445. doi:10.1016/j.tins.2010.05.006.

The molecular and gene regulatory signature of a neuron

Oliver Hobert[#], Inés Carrera, and Nikolaos Stefanakis

Columbia University Medical Center, Howard Hughes Medical Institute, New York

Abstract

Neuron-type specific gene batteries define the morphological and functional diversity of cell types in the nervous system. Here, we discuss the composition of neuron-type specific gene batteries and illustrate gene regulatory strategies employed by distinct organisms from *C.elegans* to higher vertebrates, which are instrumental in determining the unique gene expression profile and molecular composition of individual neuronal cell types. Based on principles learned from prokaryotic gene regulation, we argue that neuronal, terminal gene batteries are functionally grouped into parallel acting “regulons”. The theoretical concepts discussed here provide testable hypotheses for future experimental analysis into the exact gene regulatory mechanisms that are employed in the generation of neuronal diversity and identity.

Introduction

“Each nerve cell is a totally autonomous physiological canton”

Ramon y Cajal, 1888, formulating what became to be known as the neuron theory

“The nervous system is constituted by numerous nervous units (neurons).”

Heinrich Wilhelm Gottfried von Waldeyer, 1891, defining the word “neuron”

In 1836 and 1837 Christian Gottfried Ehrenberg and Jan Evangelista Purkinje provided the first description of nerve cells - termed “neurons” by Heinrich Wilhelm Gottfried von Waldeyer decades later [1,2]. The nature of neurons as individual and autonomous units was recognized in the 1880s through the work of the founder of modern day neuroscience, Ramon y Cajal [1]. His beautiful description of the morphology of individual neuron types revealed the basic cellular features of neurons, including dendrites, axons and what Sherrington, in 1897, described as synapses [5] (Figure 1). Following these morphological descriptions, fundamental biophysical features of neurons, such as their excitability through ionic currents and the transmission of signals at the synapse were elucidated in the first half of the twentieth century [8]. The advent of molecular neurobiology in the 1980s, a hundred years after Cajal's groundbreaking work, has put us in a position to query the molecular correlates to the morphological, biophysical and functional properties of neurons.

We will discuss here some simple theoretical concepts about how one can envision the molecular composition, i.e. the gene expression profile, of individual, mature and terminally differentiated neurons to look like and through what regulatory mechanisms such gene

[#] correspondence: or38@columbia.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

expression profiles could be brought about. We will examine primary research data to assess how these concepts bear out in reality.

The molecular signature of a neuron

Neuron type-specific identity

The differences in the morphology and function of individual mature neurons must be reflected by the differential expression of specific sets of genes. Indeed, our knowledge of gene expression patterns in the nervous system – achieved over the years through the painstaking analysis of the expression patterns of many individual genes one by one, as well as through systematic, large scale transcriptome profiling studies - has painted a picture that underscores what morphological and functional studies have already strongly indicated: In spite of some basic similarities, individual cell types in the nervous system are extraordinarily complex in their molecular composition. However, with some very rare exceptions (for example, olfactory receptors), identity-determinants are not uniquely expressed in one neuron type and nowhere else. That is, individual, mature neuron types are not defined by the expression of unique genes, but rather by a neuron type-specific combination of genes (i.e. “gene battery”; see Glossary) with each member of a battery more broadly expressed. A terminal gene battery is composed of “nuts and bolts” genes, that is, genes coding for proteins that are expressed throughout the life of the neuron and determine the shape and function of the neuron (e.g. genes coding for proteins that generate and package neurotransmitters, synaptic adhesion molecules, ion channels that shape a neuron's electrical properties, *etc.*). By our definition, such terminal batteries do not include proteins transiently expressed during development, such as axon pathfinding cues. A simple example from an individual neuron type in the nematode *C. elegans*, illustrated in a binary coding scheme, is shown in Figure 2. With the 13,000-30,000 genes that any organisms with a nervous system has at its disposal, such “combinatorial coding” allows, at least in theory, for the building of almost infinite number of combinatorial gene expression profiles. This notion goes beyond mere theory: Comparative expression profiling of two distinct neuronal cell types in the *C.elegans* has revealed more than 1,200 genes that are differentially expressed [9].

The identity of groups of neurons

Neuron types can be grouped according to specific properties. This is not a trivial exercise because, in principle, any gene expressed in multiple neuron types (see Figure 2) can be thought of as defining a group of neurons. Such grouping becomes more helpful if one doesn't consider individual genes, but if one considers functionally related “subroutines” that are shared by distinct neuron types. This is best illustrated with neurotransmitter identity. A neurotransmitter is synthesized by a set of one to several enzymes, packaged into a vesicle by specific transporters and its reuptake also usually depends on specific transporters (Figure 3). Together, the genes encoding for these proteins define a small gene battery characteristic for each neurotransmitter, executing the neuronal “subroutine” of synthesizing a specific neurotransmitter. While the neurotransmitter that a neuron utilizes is a key defining feature of a neuron, it does not uniquely define a neuron. For example, a cholinergic motor neuron in the spinal cord is vastly distinct from, say, a basal forebrain cholinergic projection neuron [13]. Other subroutines shared by distinct neurons were recently revealed through weighted analysis of gene expression patterns in the human brain, revealing expression modules that correlate with, for example, fast or slow firing properties of neurons [15].

Pan-neuronal identity

In general, all mature neurons in the nervous system – be they as diverse in appearance as they may – are alike, at least on some level. That is, all mature neurons of all organisms that contain a nervous system, no matter how primitive it is, share a number of key, “pan-neuronal” features:

(a) cellular extensions (e.g. axons and/or dendrites) which physically connect neurons and (b) synapses through which information is transmitted. Synapses themselves are made of presynaptic and postsynaptic specializations each composed of complex macromolecular assemblies (Figure 1).

Does a common molecular signature underlie these common morphological features? In other words, do all neurons express common molecular features that are shared by all neurons and that are not present in other cells? Somewhat non-intuitively, this is not the case. A survey of published expression patterns of pre-synaptic proteins (including synaptic vesicle-associated proteins), postsynaptic scaffolding proteins, ion channels and cytoskeleton proteins, shows that virtually none of these proteins are exclusively expressed in the nervous system; rather, many of them are also expressed in non-neuronal cells (see online Supplementary material, Table S1). Notably, genome sequence data accumulated from various simple organisms over the last few years has shown that unicellular organisms already contain proteins that may constitute the basic building blocks of a synapse [16-19]. With the advent of multicellularity, all cells may have initially contained this basic neuronal toolkit. Upon the specialization of cells into neurons vs. non-neuronal cells, expression of, for example, synaptic vesicle proteins then apparently become more restricted to neuronal cells, yet not completely removed from all non-neuronal cells. In summary, while neurons do share, by and large, common cellular features, there is no clear-cut “pan-neuronal gene battery”. For now, however, we will use this term to indicate very broadly expressed neuronal genes. As discussed below, such genes may be regulated by mechanisms that are distinct from those of more selectively expressed neuronal identity determinants.

Concepts for the regulatory logic of neuronal gene expression programs

Why is it useful to think about parsing neuronal identity into individual components, such as unique identity features, group-identity features and “pan-neuronal” features? Far from being a mere Linnaean exercise in grouping specimen, such concepts allow us to formulate ideas and hypotheses about a key problem that developmental neurobiologists are interested in – what are the regulatory programs that instruct the terminal identity of a neuron and how can these instructions be deciphered? These questions have traditionally been addressed using genetic approaches in which one analyzes the effect of the loss of a gene regulatory factor on the identity of a neuron type of interest. While very informative in principle, the key problem of this traditional approach has been twofold – first, in many model systems examined, the ability to precisely phenotype a gene knockout is difficult since it is often unclear what exactly happens to the neurons under examination - are they not generated, do they die, are they transformed, are they present but undifferentiated? Moreover, it is not trivial to determine at what stage a gene acts to affect the identity of a neuron type. These shortcomings help illustrate the enormous power of an alternative approach - the “bottom-up” approach in which one first defines what genes are expressed in a neuron, and then investigates the *cis*-regulatory logic through which such expression profiles are brought about (Box 1). Below, we will describe recent studies that have utilized this approach and helped further our understanding of gene regulatory strategies in the nervous system.

The regulatory logic of neuron-type specific gene expression programs

The combinatorial expression of terminal differentiation genes which define the identity of neurons (Figure 2) could be brought about in two different manners: *Selective activation* in which a terminal differentiation gene is only activated in the cells in which it is expressed; or *selective repression* in which the gene is broadly activated but selectively repressed in individual neuron types. Extensive analysis in *C. elegans*, where *cis*-regulatory control regions of individual genes can be analyzed with relative ease, has shown that unlike earlier developmental control genes, neuronal terminal differentiation genes are usually expressed

through activation-based mechanisms (e.g. [9,20-23]). In these studies, the unbiased dissection of reporter genes that monitor the expression of terminal differentiation genes expressed in individual cell types, revealed little if any evidence for repression; that is, reporter genes do not become ectopically expressed in other cell types upon removal of *cis*-regulatory regions (Box 1). The notion of “nuts and bolts” terminal differentiation determinants being regulated mainly through activation is not unique to the nervous system, but has been made for other cell types as well [24].

How then is activation in specific individual neuron types brought about? There are again two scenarios that one could consider. To be expressed in one specific neuron type, each of the many terminal differentiation genes in a neuron may be activated by a distinct set of gene regulatory factors. This is not difficult to perceive as invertebrate and vertebrate genomes contain on the order of ~1,500 transcription factors which could be expressed in a combinatorial manner in each neuron type [25]. Single neuron-transcriptome studies are, at least in theory, consistent with such a view. For example, transcriptome analysis revealed that a single gustatory neuron in *C.elegans* differentially expresses 68 transcription factors, compared to a related sensory neuron [9].

A much simpler scenario could be envisioned if one considers lessons learned in bacteria. Transcriptional regulation of functionally related bacterial genes, for example, those encoding for biosynthetic pathways that are differentially regulated in response to environmental conditions, are often co-regulated by a common transcriptional regulatory input. Batteries of co-regulated genes define what bacterial geneticists have long ago termed a “regulon” [10, 12] (Figure 4A). In modern systems biology terminology, such regulatory architecture is termed a “single input module” [26]. Surprisingly, such a simple regulatory strategy – composed of simple “regulators” and “effector genes” – may also apply to neurons. Studies in *C.elegans* have shown that different terminal gene batteries, expressed by different neurons are each co-regulated through common *cis*-regulatory motifs [9,22,23]. These common motifs are controlled by transcription factors that have been termed “terminal selectors” (Figure 4B) [6]. Loss of these particular transcription factors results in the loss of the terminal identity of individual neurons – caused by the loss of expression of terminal differentiation genes [6]. Terminal selectors are not necessarily single transcription factors, but combinations of synergistically acting transcription factors. For example, the terminal identity of the AIY interneuron class in *C.elegans*, is determined by the activity of two phylogenetically conserved homeodomain proteins, whose expression uniquely overlaps in these interneurons, and who synergistically bind to a specific *cis*-regulatory sequence (the “AIY motif”) that is found in many genes defining AIY neuron identity (Figure 2)[23,27].

The terminal selector model predicts that the *cis*-regulatory regions of terminal differentiation genes are composed of essentially linear arrays of terminal selector binding sites (“terminal selector motifs”), activating the terminal differentiation gene in the different neuron types in which the gene is expressed. Mutational analysis of *cis*-regulatory control regions of terminal differentiation genes, support this view [9,22,23] (Figure 2). There is again a striking analogy here to the bacterial regulon concept. The individual effector genes in a bacterial regulon are often not unique for a single regulon, but may be shared by different regulons that are engaged by different stimuli [28] (Figure 4C). Effector genes that are parts of distinct regulons therefore contain binding sites for distinct regulators within their *cis*-regulatory control regions. This is similar to a terminal differentiation gene (like the *sra-11* gene shown in Figure 2) that is expressed in distinct neuron types. In essence, two different regulatory states of a bacterial cell are therefore somewhat akin to two different neuron types, each expressing a partially overlapping, but distinct set of “effector genes” (Figure 4C).

Two other fundamental regulatory features are also shared between terminal selectors and bacterial regulons. One is autoregulation of the regulator (Figure 4), mediated by the presence of the regulator's own binding site in its own promoter. Autoregulation confers specific kinetic features to a system [29], but the one most relevant to the discussion here is stability of the differentiated state. That is, terminal selectors need to be persistently present so as to ensure maintenance of the differentiated state, a notion experimentally confirmed through postdevelopmental removal of terminal selectors, resulting in a loss of the differentiated state of the respective neuron type [9,22].

Another shared architectural feature of regulons and neuronal selectors is that the regulation of target genes need not be direct. The regulator may affect expression of another transcription factor, which then controls a target gene (note that sequentially acting transcription factors may be organized in a “feedforward motif”, as depicted in Figure 4B). The presence of intervening regulators means that one does not need to expect that all genes expressed in a differentiated neuron (or, in a specific environmental condition in bacteria) need to contain binding sites for the regulator. This is important to keep in mind when one undertakes a bioinformatic analysis of single cell/ state-specific transcriptomes [9].

The regulatory logic of gene expression programs that are shared between distinct groups of neurons

The neurotransmitter phenotype of a neuron is usually determined by a battery of genes that encode neurotransmitter synthetic enzymes (e.g. glutamic acid decarboxylase for GABA or tyrosine hydroxylase for dopamine), transporters that package the neurotransmitter into vesicles, and transporters that re-uptake the neurotransmitter after release (Figure 3). The amenability of *C.elegans* to both genetic knockout and transgenic reporter gene studies has shown that those neurotransmitter gene batteries are co-regulated by a common regulatory input; this is again in striking reminiscence to bacterial regulons that are involved in specific metabolic pathways. In the case of *C.elegans* GABAergic neurons, the UNC-30/Pitx2 homeodomain transcription factor controls the GABAergic phenotype through direct activation of GABAergic terminal fate markers [21] and the AST-1 ETS-domain transcription factor directly controls the dopaminergic phenotype [22] (Figure 3A). In vertebrates, the transcription factor PET-1 directly controls the serotonergic phenotype of many neurons [30] (Figure 3A). We note here that many other vertebrate transcription factors are required for the regulation of specific neurotransmitter phenotypes; however, it is not known whether this regulation is direct [6].

Does every neuron of a given neurotransmitter phenotype utilize the same transcriptional control mechanism? This appears to depend on how related the neurons are. All *C.elegans* dopaminergic neurons, even though diverse by lineage history, are functionally related – they are mechanosensory neurons. And in this case, AST-1 indeed controls the dopaminergic phenotype in *all* dopaminergic neurons [22]. *C.elegans* GABAergic neurons, in contrast, are functionally and morphologically more diverse and UNC-30 controls the GABAergic phenotype only in a group of functionally related ventral nerve cord motor neurons, but not in other interneurons or motorneurons in the worm head [31]. Importantly, both UNC-30 and AST-1 appear to be terminal selector genes – that is, they not only control neurotransmitter phenotype but all other unique features that define the respective neuronal types [22,32]. Neurotransmitter identity is therefore one of several integral parts of a co-regulated gene battery that defines neuronal identity.

Another example of features shared by many distinct neurons are genes that are expressed in all different types of sensory neurons and that encode basic components of the ciliated specializations of sensory dendrites. These genes constitute a “pan-sensory battery”. Cilia structures appear to be composed of two different components – basic components shared by

every ciliated neuron (e.g. specific kinesins or intraflagellar transport core subunits) and more specialized ciliated building blocks that make cilia look and work differentially in different types of sensory neurons [33]. Basic components of cilia, i.e. the pan-sensory gene battery, are indeed co-regulated by a common transcriptional input, again akin to classic bacterial regulons (Figure 3B). This co-regulation happens in all sensory neurons. That is, pan-sensory genes share a common *cis*-regulatory motif, the X-box, which is required for their expression in all sensory neurons and which constitutes a binding site for the RFX-type DAF-19 transcription factor [34,35]. Consequently, genetic elimination of *daf-19* results in a loss of ciliated structures. This regulatory mechanism operates in different types of sensory neurons and, importantly, is also independent of terminal selector genes because elimination of terminal selectors does not affect the ciliated sensory identity of neurons [22,36]. Notably, the function of DAF-19 appears to also be conserved in flies and vertebrates [33].

Taken together, the regulation of group identity features is diverse. Some appear to be regulated in a piece-meal manner (i.e. different transcription factors in different neuron types), whereas others are “bracketed” by common regulatory mechanism that operate in distinct neuron types.

The regulatory logic of pan-neuronal gene expression

Broadly expressed, pan-neuronal genes appear to be regulated in a manner distinct from that of more specifically expressed, neuronal identity determinants. This notion is based on the observation that the loss of terminal selector genes by and large does not affect pan-neuronal features of the neuron. This has been observed in several different cases, thereby making the point that a neuron's individuality is genetically separable from its overall identity as a neuron [22,27,36]. The other reason to separate the regulatory logic of “pan-neuronal” genes from that of more specific identity determinants, is based on bioinformatic analysis which revealed a sequence motif in the regulatory region of many broadly expressed pan-neuronal genes in *C.elegans*, termed the “N1-box”, [37]. The presence of such a shared motif suggests that an as yet unidentified *trans*-acting factor exists, which is dedicated to bringing about pan-neuronal features. Such shared *cis*-regulatory motif(s) are not restricted to *C.elegans* but also occur in pan-neuronal genes in the chordate *Ciona intestinalis* [38]. In vertebrates, shared *cis*-regulatory motifs have been found in the regulatory region of genes coding for pre-synaptic proteins [39]. These bioinformatic studies, together with the mutational analysis in worms suggest, at least at first sight, that pan-neuronally expressed genes are activated by a common, broadly expressed pan-neuronal activator. Moreover, they support the notion that many pan-neuronal features are - like other subroutines in a neuron - co-regulated by a common transcriptional regulatory input.

While bioinformatic analysis has pointed to the existence of some broad, co-regulatory activation mechanisms, such analysis has also made the point that this is not the entire story. For example, the N1 box is clearly not present in a subset of bona-fide pan-neuronal genes in *C. elegans*. For those genes, the alternative mechanism to get broad neuronal activation appears to be a “piece-meal” activation strategy, mediated by an additive array of binding sites for terminal selector genes that result in broad neuronal activation. The evidence for the “piece-meal” mechanism originates again from *cis*-regulatory analysis. Mutational analysis of broad neuronal genes, such as β 1-tubulin in flies [40] or SNAP-25 in worms [41] reveals that the *cis*-regulatory control regions of these pan-neuronal genes are composed of separable *cis*-regulatory elements, each responsible for gene activation in neuronal subsets. In the case of SNAP-25, one of these *cis*-regulatory control regions contains a functional binding site for the MEC-3/UNC-86 terminal selector complex [41].

Taken together, regulatory strategies for achieving broad or pan-neuronal gene expression patterns appear to be surprisingly diverse. Some pan-neuronal genes may be controlled in a piece-meal manner by sets of terminal selector genes, while for others, broad activators may

exist that act in parallel to terminal selectors (Figure 5). Combinations of both scenarios can also be envisioned. Clearly, the available data is still somewhat anecdotal and it is imperative to conduct experiments that more systematically and comprehensively probe the regulatory strategies that promote such pan-neuronal gene expression.

Regulatory principles in invertebrate versus vertebrate nervous systems

Decades of molecular genetic analysis in vertebrate and invertebrate model systems have revealed conservation of many basic principle of nervous system development and function. A classic example for such conservation is the proneural activity of basic helix-loop-helix transcription factors [42] and a more recent example is the apparent conservation of the regulatory logic of dopaminergic neuron differentiation by ETS domain transcription factors [22]. Core regulatory principles, like autoregulation of transcription factors that drive neuronal differentiation, combinatorial activity of transcription factors and the organization of transcription factor activities into network motifs like feedforward circuits are also broadly conserved throughout evolution (e.g. [15,43-45]). The organization of functionally related proteins into regulatory modules (“regulons”) has also begun to be uncovered through microarray-based weighted gene co-expression network analysis [15]. Candidate terminal selectors that act through shared *cis*-regulatory motifs have been identified in mammalian serotonergic neurons and in mammalian photoreceptor neurons [45,46].

Yet the advent of more complex organisms may also have been accompanied by the invention of novel molecules and strategies allowing for the generation of further diversity. A case in point for such a new invention may be the zinc finger transcription factor, REST(RE1-silencing transcription factor)/NRSF(neuron restricted silencing factor) [47,48], found in vertebrates, but not in flies, worms, sea urchin or sea squirts. REST/NRSF is expressed in many (but not all) non-neuronal cells and through binding to its cognate binding site (RE1/NRSE) represses the expression of terminal neuronal differentiation genes in non-neuronal tissues [47-52]. REST/NRSF targets many terminal differentiation genes, some very broadly expressed, some more neuron-type specifically expressed. REST/NRSF-mediated control of gene expression does not challenge, but significantly extends some of the principles discussed above. In particular, many of the genes targeted by REST/NRSF appear to be selectively expressed in only some neuron types and this selective expression appears to be mediated by neuron-type specific transcription factors (e.g. [53,54]). REST/NRSF therefore refines the cell and tissue-selectivity of gene expression profiles that were set up by specific gene activation events. Consistent with the notion of this being an “addition” to a more ancestral gene activation principle, target genes of REST/NRSF are broadly conserved across phylogeny, yet REST/NRSF clearly is not. Moreover, note that extensive *cis*-regulatory analysis of terminal differentiation genes in invertebrates has not yet provided much support for the existence of repressive regulatory mechanisms, as described above and in Box 1.

Summary and concluding remarks

Even though the morphological diversity of neuronal cell types has been recognized over a century ago, we are still only beginning to appreciate the underlying molecular diversity of neuron types. One key limitation has been the technical difficulty of reliably defining, isolating in sufficient quantity, and molecularly profiling individual neurons from a complex nervous system. Substantial progress in this area, both in vertebrate and invertebrate neurons, has been achieved in the last few years [9,32,55-57]. This has surely led to an even more comprehensive appreciation of the molecular complexity of a terminally differentiated neuron. Yet it remains striking how little we know about the regulation of those “nuts and bolts” genes that are key determinants of the functional property of a neuron. Systematic, non-anecdotal, *in vivo* analyses of the *cis*-regulatory control region of individual terminal differentiation markers on a single

neuron level has so far essentially only been performed in the *C.elegans* nervous system, a system particularly amenable to comprehensive reporter gene studies. These types of *cis*-regulatory analyses, as well as weighted co-expression network analysis in vertebrates [15, 58], are beginning to provide testable hypotheses and concepts about how one can subdivide the nature of neuronal gene batteries, and about what types of regulatory strategies may be employed to control these gene expression programs.

One emerging regulatory concept is the co-regulation of terminal differentiation genes through shared *cis*-regulatory motifs and cognate transcription factors, a concept much akin to the long-recognized regulon concept in bacteria. Like in bacteria, this provides a relatively flat regulatory hierarchy, grouping together functionally related genes (i.e. genes that control the terminal phenotype of a vertebrate neuron, or genes that respond to a specific environmental condition in bacteria). Not every terminal feature is controlled by a single regulon; rather, there are almost certainly several of them acting in parallel, each regulating distinct aspects of the terminally differentiated state of a neuron. This concept is not self-evident. The number of transcription factors in the genome, and the number that are expressed in a neuron, surely leave room to think about a much more “messy”, piece-meal nature of gene regulation. Such messiness will surely exist to some extent, but the terminal selector/ regulon concept may constitute an ancient “pattern” that may have simply become more and more diversified, through the addition of more and more transcription factors that execute additional subroutines.

Clearly, more experimental analysis is required to carve out common themes in the regulatory programs that determine the identity of a mature neuron, the fundamental unit of the nervous system. The importance of such an understanding goes much beyond basic biology. For instance, a better understanding of how to precisely drive a specific differentiation program to derive unique cell types may be a worthy strategy for cellular replacement therapies for various human pathological disorders.

Box 1

Analyzing *cis*-regulatory control of gene expression patterns in transgenic animals

This schematic illustrates the approach of defining *cis*-regulatory control elements in transgenic animals. Genomic regions adjacent to the gene of interest are fused to a fluorescent reporter and the DNA construct is injected into animals (in this example, *C.elegans*) to generate transgenic animals that will reveal the *cis*-regulatory content of each piece of DNA. The example shown here is from the *cis*-regulatory analysis of a gene, the tyramine receptor-encoding *ser-2* gene, in the *C.elegans* AIY interneuron [23]; schematic results of this analysis are also shown in Figure 1. For example, only those constructs that contain the AIY terminal selector motif of the *ser-2* gene allow expression in AIY interneurons, whereas other sequence motifs specific *ser-2* expression in other neuronal cells (e.g. RID, OLL, DA9) and head muscle cells. Unlike earlier studies of *cis*-regulatory analyses conducted in vertebrate cell lines (e.g. [70]), this approach probes the function of *cis*-regulatory elements within the whole animal, hence within the cellular context in which the gene of study is normally expressed, and where the correct complement of trans-acting factors exists. As detailed in the text, such analyses have revealed three fundamental principles of gene regulation of terminal neuronal genes; first, regulation is modular with individual elements driving expression in different cell types; second, more fine-grained mutational analysis showed that expression in individual cell types depends on single *cis*-regulatory motifs (here: the AIY motif, a binding site for the TTX-3/CEH-10 terminal selector complex; Figure 1); third, little evidence for repression playing a role in shaping expression patterns was found. If repression were to play a major role, some of the reporters should have revealed derepression in other cell types, which was only rarely observed.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Work in our laboratory is funded by the National Institute of Health (R01NS039996-05; R01NS050266-03) and the Howard Hughes Medical Institute. We thank Piali Sengupta, Richard Goodman, Gail Mandel, Richard Mann and members of the Hobert lab for comments on the manuscript.

References

1. Lopez-Munoz F, Boya J, Alamo C. Neuron theory, the cornerstone of neuroscience, on the centenary of the Nobel Prize award to Santiago Ramon y Cajal. *Brain Res Bull* 2006;70:391–405. [PubMed: 17027775]
2. Waldeyer W. Ueber einige neuere Forschungen im Gebiete der Anatomie des Centralnervensystems. *Deutsche Medicinische Wochenschrift* 1891;50:1352–1356.
3. Garcia-Bellido A. Genetic control of wing disc development in *Drosophila*. *Ciba Found Symp* 1975;29:161–182. [PubMed: 1039909]
4. Mann RS, Carroll SB. Molecular mechanisms of selector gene function and evolution. *Curr Opin Genet Dev* 2002;12:592–600. [PubMed: 12200165]
5. Tansey EM. Not committing barbarisms: Sherrington and the synapse, 1897. *Brain Res Bull* 1997;44:211–212. [PubMed: 9323432]
6. Hobert O. Regulatory logic of neuronal diversity: terminal selector genes and selector motifs. *Proc Natl Acad Sci U S A* 2008;105:20067–20071. [PubMed: 19104055]
7. Morgan, TH. *Embryology and Genetics*. New York: Columbia University Press; 1934.

8. Albright TD, Jessell TM, Kandel ER, Posner MI. Neural science: a century of progress and the mysteries that remain. *Cell* 2000;100(Suppl):S1–55. [PubMed: 10698132]
9. Etchberger JF, Lorch A, Sleumer MC, Zapf R, Jones SJ, Marra MA, Holt RA, Moerman DG, Hobert O. The molecular signature and cis-regulatory architecture of a *C. elegans* gustatory neuron. *Genes Dev* 2007;21:1653–1674. [PubMed: 17606643]
10. Maas WK, Clark AJ. Studies on the mechanism of repression of arginine biosynthesis in *Escherichia coli*: II. Dominance of repressibility in diploids. *J Mol Biol* 1964;8:365–370. [PubMed: 14168690]
11. Cajal, R. *Histology*. 10th. Baltimore: Wood; 1933.
12. Epstein W, Beckwith J. Regulation of Gene Expression. *Annu Rev Biochem* 1968;37:411–436.
13. Woolf NJ. Cholinergic systems in mammalian brain and spinal cord. *Prog Neurobiol* 1991;37:475–524. [PubMed: 1763188]
14. Yoshida K, Yamaguchi H, Kinehara M, Ohki YH, Nakaura Y, Fujita Y. Identification of additional TnrA-regulated genes of *Bacillus subtilis* associated with a TnrA box. *Mol Microbiol* 2003;49:157–165. [PubMed: 12823818]
15. Winden KD, Oldham MC, Mirmics K, Ebert PJ, Swan CH, Levitt P, Rubenstein JL, Horvath S, Geschwind DH. The organization of the transcriptional network in specific neuronal classes. *Mol Syst Biol* 2009;5:291. [PubMed: 19638972]
16. Putnam NH, Srivastava M, Hellsten U, Dirks B, Chapman J, Salamov A, Terry A, Shapiro H, Lindquist E, Kapitonov VV, et al. Sea anemone genome reveals ancestral eumetazoan gene repertoire and genomic organization. *Science* 2007;317:86–94. [PubMed: 17615350]
17. Kosik KS. Exploring the early origins of the synapse by comparative genomics. *Biol Lett* 2009;5:108–111. [PubMed: 19049956]
18. Ryan TJ, Grant SG. The origin and evolution of synapses. *Nat Rev Neurosci* 2009;10:701–712. [PubMed: 19738623]
19. Sakarya O, Armstrong KA, Adamska M, Adamski M, Wang IF, Tidor B, Degnan BM, Oakley TH, Kosik KS. A post-synaptic scaffold at the origin of the animal kingdom. *PLoS ONE* 2007;2:e506. [PubMed: 17551586]
20. Nokes EB, Van Der Linden AM, Winslow C, Mukhopadhyay S, Ma K, Sengupta P. Cis-regulatory mechanisms of gene expression in an olfactory neuron type in *Caenorhabditis elegans*. *Dev Dyn* 2009;238:3080–3092. [PubMed: 19924784]
21. Eastman C, Horvitz HR, Jin Y. Coordinated transcriptional regulation of the unc-25 glutamic acid decarboxylase and the unc-47 GABA vesicular transporter by the *Caenorhabditis elegans* UNC-30 homeodomain protein. *J Neurosci* 1999;19:6225–6234. [PubMed: 10414952]
22. Flames N, Hobert O. Gene regulatory logic of dopamine neuron differentiation. *Nature* 2009;458:885–889. [PubMed: 19287374]
23. Wenick AS, Hobert O. Genomic cis-Regulatory Architecture and trans-Acting Regulators of a Single Interneuron-Specific Gene Battery in *C. elegans*. *Dev Cell* 2004;6:757–770. [PubMed: 15177025]
24. Arnone MI, Davidson EH. The hardwiring of development: organization and function of genomic regulatory systems. *Development* 1997;124:1851–1864. [PubMed: 9169833]
25. Gray PA, Fu H, Luo P, Zhao Q, Yu J, Ferrari A, Tenzen T, Yuk DI, Tsung EF, Cai Z, et al. Mouse brain organization revealed through direct genome-scale TF expression analysis. *Science* 2004;306:2255–2257. [PubMed: 15618518]
26. Alon U. Network motifs: theory and experimental approaches. *Nat Rev Genet* 2007;8:450–461. [PubMed: 17510665]
27. Altun-Gultekin Z, Andachi Y, Tsalik EL, Pilgrim D, Kohara Y, Hobert O. A regulatory cascade of three homeobox genes, *ceh-10*, *ttx-3* and *ceh-23*, controls cell fate specification of a defined interneuron class in *C. elegans*. *Development* 2001;128:1951–1969. [PubMed: 11493519]
28. Neidhardt, FC.; Savageau, MA. Regulation beyond the Operon In *Escherichia coli* and *Salmonella*: Cellular and molecular biology. 2nd. Neidhardt, FC.; Curtis, R.; Ingraham, JL.; Lin, ECC.; Low, KB.; Magasanik, B.; Reznikoff, WS.; Riley, M.; Schaechter, M.; Umberger, HE., editors. Vol. 1. Washington, D.C.: ASM Press; 1996. p. 1310-1324.
29. Alon, U. *An Introduction to Systems Biology: Design Principles of Biological Circuits*. 1st. Chapman & Hall/CRC; 2006.

30. Hendricks T, Francis N, Fyodorov D, Deneris ES. The ETS domain factor Pet-1 is an early and precise marker of central serotonin neurons and interacts with a conserved element in serotonergic genes. *J Neurosci* 1999;19:10348–10356. [PubMed: 10575032]
31. Jin Y, Hoskins R, Horvitz HR. Control of type-D GABAergic neuron differentiation by *C. elegans* UNC-30 homeodomain protein. *Nature* 1994;372:780–783. [PubMed: 7997265]
32. Cinar H, Keles S, Jin Y. Expression profiling of GABAergic motor neurons in *Caenorhabditis elegans*. *Curr Biol* 2005;15:340–346. [PubMed: 15723795]
33. Inglis PN, Ou G, Leroux MR, Scholey JM. The sensory cilia of *Caenorhabditis elegans*. *WormBook* 2007:1–22. [PubMed: 18050505]
34. Swoboda P, Adler HT, Thomas JH. The RFX-type transcription factor DAF-19 regulates sensory neuron cilium formation in *C. elegans*. *Mol Cell* 2000;5:411–421. [PubMed: 10882127]
35. Efimenko E, Bubb K, Mak HY, Holzman T, Leroux MR, Ruvkun G, Thomas JH, Swoboda P. Analysis of *xbx* genes in *C. elegans*. *Development* 2005;132:1923–1934. [PubMed: 15790967]
36. Uchida O, Nakano H, Koga M, Ohshima Y. The *C. elegans* *che-1* gene encodes a zinc finger transcription factor required for specification of the ASE chemosensory neurons. *Development* 2003;130:1215–1224. [PubMed: 12588839]
37. Ruvinsky I, Ohler U, Burge CB, Ruvkun G. Detection of broadly expressed neuronal genes in *C. elegans*. *Dev Biol* 2007;302:617–626. [PubMed: 17046742]
38. Kusakabe T, Yoshida R, Ikeda Y, Tsuda M. Computational discovery of DNA motifs associated with cell type-specific gene expression in *Ciona*. *Dev Biol* 2004;276:563–580. [PubMed: 15581886]
39. Liu R, Hannehalli S, Bucan M. Motifs and cis-regulatory modules mediating the expression of genes co-expressed in presynaptic neurons. *Genome Biol* 2009;10:R72. [PubMed: 19570198]
40. Kohler J, Schafer-Preuss S, Buttgerit D. Related enhancers in the intron of the beta1 tubulin gene of *Drosophila melanogaster* are essential for maternal and CNS-specific expression during embryogenesis. *Nucleic Acids Res* 1996;24:2543–2550. [PubMed: 8692694]
41. Hwang SB, Lee J. Neuron cell type-specific SNAP-25 expression driven by multiple regulatory elements in the nematode *Caenorhabditis elegans*. *J Mol Biol* 2003;333:237–247. [PubMed: 14529613]
42. Bertrand N, Castro DS, Guillemot F. Proneural genes and the specification of neural cell types. *Nat Rev Neurosci* 2002;3:517–530. [PubMed: 12094208]
43. Edlund T, Jessell TM. Progression from extrinsic to intrinsic signaling in cell fate specification: a view from the nervous system. *Cell* 1999;96:211–224. [PubMed: 9988216]
44. Baumgardt M, Miguel-Aliaga I, Karlsson D, Ekman H, Thor S. Specification of neuronal identities by feedforward combinatorial coding. *PLoS Biol* 2007;5:e37. [PubMed: 17298176]
45. Hsiao TH, Diaconu C, Myers CA, Lee J, Cepko CL, Corbo JC. The cis-regulatory logic of the mammalian photoreceptor transcriptional network. *PLoS ONE* 2007;2:e643. [PubMed: 17653270]
46. Hendricks TJ, Fyodorov DV, Wegman LJ, Lelutiu NB, Pehek EA, Yamamoto B, Silver J, Weeber EJ, Sweatt JD, Deneris ES. Pet-1 ETS gene plays a critical role in 5-HT neuron development and is required for normal anxiety-like and aggressive behavior. *Neuron* 2003;37:233–247. [PubMed: 12546819]
47. Schoenherr CJ, Anderson DJ. The neuron-restrictive silencer factor (NRSF): a coordinate repressor of multiple neuron-specific genes. *Science* 1995;267:1360–1363. [PubMed: 7871435]
48. Chong JA, Tapia-Ramirez J, Kim S, Toledo-Aral JJ, Zheng Y, Boutros MC, Altschuller YM, Frohman MA, Kraner SD, Mandel G. REST: a mammalian silencer protein that restricts sodium channel gene expression to neurons. *Cell* 1995;80:949–957. [PubMed: 7697725]
49. Otto SJ, McCorkle SR, Hover J, Conaco C, Han JJ, Impey S, Yochum GS, Dunn JJ, Goodman RH, Mandel G. A new binding motif for the transcriptional repressor REST uncovers large gene networks devoted to neuronal functions. *J Neurosci* 2007;27:6729–6739. [PubMed: 17581960]
50. Schoenherr CJ, Paquette AJ, Anderson DJ. Identification of potential target genes for the neuron-restrictive silencer factor. *Proc Natl Acad Sci U S A* 1996;93:9881–9886. [PubMed: 8790425]
51. Chen ZF, Paquette AJ, Anderson DJ. NRSF/REST is required in vivo for repression of multiple neuronal target genes during embryogenesis. *Nat Genet* 1998;20:136–142. [PubMed: 9771705]

52. Johnson DS, Mortazavi A, Myers RM, Wold B. Genome-wide mapping of in vivo protein-DNA interactions. *Science* 2007;316:1497–1502. [PubMed: 17540862]
53. Mu X, Fu X, Sun H, Beremand PD, Thomas TL, Klein WH. A gene network downstream of transcription factor Math5 regulates retinal progenitor cell competence and ganglion cell fate. *Dev Biol* 2005;280:467–481. [PubMed: 15882586]
54. Bessis A, Salmon AM, Zoli M, Le Novère N, Picciotto M, Changeux JP. Promoter elements conferring neuron-specific expression of the beta 2-subunit of the neuronal nicotinic acetylcholine receptor studied in vitro and in transgenic mice. *Neuroscience* 1995;69:807–819. [PubMed: 8596650]
55. Colosimo ME, Brown A, Mukhopadhyay S, Gabel C, Lanjuin AE, Samuel AD, Sengupta P. Identification of thermosensory and olfactory neuron-specific genes via expression profiling of single neuron types. *Curr Biol* 2004;14:2245–2251. [PubMed: 15620651]
56. Fox RM, Von Stetina SE, Barlow SJ, Shaffer C, Olszewski KL, Moore JH, Dupuy D, Vidal M, Miller DM 3rd. A gene expression fingerprint of *C. elegans* embryonic motor neurons. *BMC Genomics* 2005;6:42. [PubMed: 15780142]
57. Nelson SB, Hempel C, Sugino K. Probing the transcriptome of neuronal cell types. *Curr Opin Neurobiol* 2006;16:571–576. [PubMed: 16962313]
58. Oldham MC, Konopka G, Iwamoto K, Langfelder P, Kato T, Horvath S, Geschwind DH. Functional organization of the transcriptome in human brain. *Nat Neurosci* 2008;11:1271–1282. [PubMed: 18849986]
59. Kandel, ER.; Jessell, TM.; Schwartz, JM. Principles of Neural Science. Appleton & Lange; 1991.
60. Feng W, Zhang M. Organization and dynamics of PDZ-domain-related supramodules in the postsynaptic density. *Nat Rev Neurosci* 2009;10:87–99. [PubMed: 19153575]
61. Garner CC, Kindler S, Gundelfinger ED. Molecular determinants of presynaptic active zones. *Curr Opin Neurobiol* 2000;10:321–327. [PubMed: 10851173]
62. Olsen O, Moore KA, Nicoll RA, Brecht DS. Synaptic transmission regulated by a presynaptic MALS/Liprin-alpha protein complex. *Curr Opin Cell Biol* 2006;18:223–227. [PubMed: 16504495]
63. Hobert O. Specification of the nervous system. *WormBook* 2005:1–19. [PubMed: 18050401]
64. Robichon D, Arnaud M, Gardan R, Pragai Z, O'Reilly M, Rapoport G, Debarbouille M. Expression of a new operon from *Bacillus subtilis*, ykzB-ykoL, under the control of the TnrA and PhoP-phoR global regulators. *J Bacteriol* 2000;182:1226–1231. [PubMed: 10671441]
65. Brandenburg JL, Wray LV Jr, Beier L, Jarmer H, Saxild HH, Fisher SH. Roles of PucR, GlnR, and TnrA in regulating expression of the *Bacillus subtilis* ure P3 promoter. *J Bacteriol* 2002;184:6060–6064. [PubMed: 12374841]
66. Davidson EH, McClay DR, Hood L. Regulatory gene networks and the properties of the developmental process. *Proc Natl Acad Sci U S A* 2003;100:1475–1480. [PubMed: 12578984]
67. Johnston RJ Jr, Copeland JW, Fasnacht M, Etchberger JF, Liu J, Honig B, Hobert O. An unusual Zn-finger/FH2 domain protein controls a left/right asymmetric neuronal fate decision in *C. elegans*. *Development* 2006;133:3317–3328. [PubMed: 16887832]
68. Etchberger JF, Flowers EB, Poole RJ, Bashllari E, Hobert O. Cis-regulatory mechanisms of left/right asymmetric neuron-subtype specification in *C. elegans*. *Development* 2009;136:147–160. [PubMed: 19060335]
69. Impey S, McCorkle SR, Cha-Molstad H, Dwyer JM, Yochum GS, Boss JM, McWeeney S, Dunn JJ, Mandel G, Goodman RH. Defining the CREB regulon: a genome-wide analysis of transcription factor regulatory regions. *Cell* 2004;119:1041–1054. [PubMed: 15620361]
70. Mori N, Stein R, Sigmund O, Anderson DJ. A cell type-preferred silencer element that controls the neural-specific expression of the SCG10 gene. *Neuron* 1990;4:583–594. [PubMed: 2322462]

Glossary

Transcriptional co-regulation	A term that indicates that the transcription of a gene battery is controlled by the same <i>cis</i> -regulatory element and cognate transcription factor(s).
-------------------------------	--

Terminal differentiation gene	Genes expressed in a mature, non-dividing neuron throughout the life of a neuron and determining its functional properties, such as ion channels, neurotransmitter-synthesizing enzymes, etc. We also refer to them here as the “nuts and bolts” genes because their gene products define the workings of an individual neuron.
Selector gene	A term proposed by Garcia-Bellido [3] to describe genes - uncloned at that time - that control the identity of a developing field, compartment or organ. See also Ref. [4].
Neuronal terminal selector gene	Extending the selector gene term, these genes code for transcription factors that co-regulate the battery of individual terminal differentiation genes that define individual neuron types through direct binding to the <i>cis</i> -regulatory control regions of these genes [6]. Genetic elimination of a neuronal terminal selector results in a differentiation defect and identity loss of the respective neuron. Terminal selectors can be hetero-/multimeric transcription factor complexes.
Pan	From Greek for “all”. Used here to note broad expression in a large group of cells, e.g. pan-neuronal or pan-sensory.
Gene battery	A term first coined by Thomas Hunt Morgan [7]. The modern day meaning is a specific set of genes that together define a specific property, e.g. a gene battery that defines a specific cell type or a specific neurotransmitter phenotype or is induced upon a specific environmental condition. The gene batteries discussed in this paper are mostly “terminal gene batteries”, i.e. batteries of terminal differentiation genes.
Combinatorial codes	The development and terminal identity of a neuron are defined by combinatorial gene expression profiles. Gene expression profiles in a cell are combinatorial in the sense that it is the combination of specific genes rather than single genes that uniquely define the development and identity of an individual neuron type. Combinatorial expression of terminal differentiation genes uniquely define, and thereby encode the terminal properties of a neuron, while combinatorial expression of transcription factor-encoding genes uniquely define the regulatory state of a cell.
Gene regulatory factor	A <i>trans</i> -acting factor with sequence-specific nucleic acid binding activity. In this review, we discuss the most prominent class, DNA binding transcription factors. Other factors, such as microRNAs, are discussed elsewhere [6].
Regulon	A term from bacterial genetics, in which a battery of genes (“effector genes”), organized into several spatially distinct operons, show coordinated expression [10-12]. Subsequent molecular analysis revealed “global regulators”, i.e. transcription factors that directly bind to the <i>cis</i> -regulatory region of genes within a regulon (e.g. [14]). Global regulators are the bacterial analog of terminal selector genes.
Subroutine	In computer science, a subroutine is a portion of code within a larger program, performing a specific task. We use this term here to designate groups of functionally related genes that perform a specific task within a neuron, e.g. a battery of genes that code for proteins that synthesize, package and re-uptake a neurotransmitter.

Fig.1. Common features of neurons

Schematic representations of three neuronal subtypes (spinal cord motor neuron, hippocampal pyramidal neuron and cerebellar Purkinje cell) with “generic” components of the pre- and postsynaptic apparatus highlighted. Proteins labeled in red are generally considered pan-neuronal (one argument, for example, being that they already are present in organisms that contain “proto-synapses”), while proteins in yellow are cell-type specific. For more details on these generic pre- and postsynaptic proteins, see [17,19,60-62] and Suppl. Table 1.

Fig.2. Terminal differentiation programs and the concept of combinatorial coding

The identity of a neuron is not determined by the expression of unique genes, but the specific combination of genes which themselves may be more broadly expressed. One example is shown here, yet there are many more examples from different invertebrate and vertebrate neuron types (e.g.[57]). Even with just the 5 genes shown (the *sra-11* serpentine receptor, the secreted factors *kal-1* and *hen-1*, the acetylcholine transporter *unc-17* and the tyramine receptor *ser-2*), a theoretical diversity of $2^5 (= 32)$ cell specific batteries could be generated; yet, single neuron transcriptome studies have shown that even functionally related neurons express >1,000 genes in a differential manner [9]. Combinatorial coding is not just involved in defining specific terminal features of a neuron, but also brings about different transcription factor combinations in individual cell types. In the example shown, the two homeodomain proteins, TTX-3 and CEH-10, synergistically bind to a *cis*-regulatory element, the AIY motif (red circle), found in terminal differentiation genes that define the terminal identity of the AIY neuron type [23]. Note that individual genes, such as the G-protein coupled receptor-encoding (*sra-11*) gene shown here, are activated in other neuron types (e.g. the AIA neuron type and other neuronal cells) through separable regulatory elements (black and grey circles, respectively), which may

constitute binding sites for other terminal selector genes. Other neurons types (egg. RID, CAN, ADL, AIZ, RMED, SIA and ASE) do not express the *sra-11* gene. Components of the genetic signatures of individual neuron types are not restricted to the nervous system. In the example shown here, the *ser-2* gene is also expressed in muscle cells. Adapted, with permission, from Ref. [63].

Fig.3. Neuronal subroutines are shared by many different neuron types

Two examples of group identities, shared by distinct neuron types are shown. These subroutines are controlled by a co-regulatory mechanism in which one transcription factor (or a combination of several transcription factors) control each individual feature.

(A) Neurotransmitter identity is co-regulated. AST-1 (an ETS domain transcription factor) and UNC-30 (a homeodomain transcription factor) are examples of transcription factors that regulate the neurotransmitter synthesis machinery and uptake transporters in the presynaptic terminal of dopaminergic synapses and motor neuron GABAergic synapses of *C.elegans*, respectively [21,22]. PET-1 (an ETS domain transcription factor) is an example of a transcription factor that controls the synthesis and reuptake machinery for serotonin (5-HT) in serotonergic synapses of mice [30]. Note that in the case of AST-1 and UNC-30, there are also other likely, as yet unknown specificity determinants of the neurotransmitter identity of these terminals, as those particular transcription factors are expressed more broadly than in just dopaminergic and GABAergic neurons, respectively.

(B) Ciliated features of sensory neurons are co-regulated. Core features of cilia, such as kinesin motor proteins, the intraflagellar transport (IFT) complex, Bardet-Biedl syndrome (BBS)

proteins, and cilia specific proteins, are regulated in all types of different ciliated sensory neurons by the phylogenetically conserved, RFX-type transcription factor, DAF-19 [34]. Sensory neuron-specific variations in ciliated structure are controlled by other transcription factors working possibly in combination with DAF-19.

Fig.4. Regulons and the terminal selector concept

(A) Bacterial regulons are constituted by functionally related “effector genes” that bacteria use to respond to specific stimuli [10]. The “effectors” within a regulon are controlled by “global regulators”, i.e. transcription factors that activate (or sometimes repress) the effector genes. As an example, the structure of the regulon involved in nitrogen metabolism in the soil microbe *Bacillus subtilis*, is shown (not all targets of the TnrA transcription factor are shown; also note that effector genes are usually organized into operons, as shown here) [14,64,65].

(B) Key features of the structure of a bacterial regulon are preserved in terminal selector genes of postmitotic neurons. Functionally related genes, i.e., the terminal gene battery of a neuron (analogous to the “effector” genes in bacteria), are co-regulated through shared *cis*-regulatory motifs and a common cognate transcription factor. Note that terminal selectors need not be single proteins, but combinations of transcription factors that may act together; for example TTX-3 and CEH-10, as shown in Figure 2). The architecture of gene regulation through terminal selectors encompass specific network motifs which are prominently found in many distinct systems (e.g. “autoregulatory loops”, “feedforward motifs” and “single input motifs”)

[26]. Autoregulation is a feature shared by many transcription factors with a role in development [66]. Feedforward loops (composed of a terminal selector, a transcription factor (TF) target and a terminal target gene) have been characterized in terminal differentiation programs of *C.elegans* gustatory neurons and interneurons [27,67,68], postmitotic *Drosophila* neurons [44] as well as mammalian photoreceptors [45] and may endow gene regulation with specific kinetic properties [29]. For example, the terminal selector CHE-1 directly activates *ceh-36*, a homeobox gene. Both CHE-1 and CEH-36 then bind to the *cis*-regulatory region of the *gcy-7* gene, which encodes a putative chemoreceptor [68]. Other TFs that are targets of terminal selectors may act alone to regulate the expression of small regulatory subroutines [9].

(C) Regulatory states of cells are defined by overlapping sets of genes. Gene expression programs in distinct neuronal cell types, or in bacterial cells under specific conditions, may operate by similar principles in the sense that (i) cell/ environmental state-defining genes are co-regulated and (ii) some of the genes may be shared by distinct cell types/ environmental states. Note that the some effector genes [e.g., X_5 and X_6 , which are shared by several regulatory states (or cell types)], contain two binding sites for their upstream regulators in their *cis*-regulatory control region (as shown schematically also in Figure 2 for the *sra-11* gene).

Fig.5. The abundance of regulons - a hypothesis

Terminal gene batteries of a neuron type may be characterized by a series of parallel acting regulons. The number and composition of regulons may differ in distinct neuron types, as illustrated in the examples shown in A-C. Circles indicate shared *cis*-regulatory motifs.

(A) In the case of the ASE gustatory neurons, an extensively characterized *C. elegans* neuron type, a clear separation between the individual regulatory routines can be observed [9]. A terminal selector, the CHE-1 zinc finger transcription factor, appears to regulate neuron-identity defining features of ASE (see also Figure 4B), but not its pan-neuronal features (which are regulated by as yet unknown means), nor its pan-sensory features (which are regulated by the RFX-type transcription factor DAF-19, see Figure 3B).

(B). In other neuron types, the separation between regulatory programs may be less tight in that pan-neuronal features may be controlled by regulators that control neuron-identity features (e.g. *C.elegans* UNC-86/MEC-3 regulating SNAP-25 [41]).

(C) In some neurons, there may not be one selector gene for all identity features, but several ones (e.g. indicated as red and yellow gene differentiation programs) that regulate each distinct part of the neuron's specific identity. This notion is also supported by recent weighted gene co-expression network analysis in complex nervous systems, which revealed the existence of various distinct, co-regulated subroutines of gene expression programs that define individual aspects of a neuronal identity [15]. In addition to “hard-wired” regulatory programs that reproducibly and invariably define individual neuron types, neurons of course also respond to specific stimuli by adapting and changing gene expression programs, and these stimulus-dependent gene expression programs are also likely to be organized into regulons, with one example being the cAMP-regulated CREB regulon [69].