PERSPECTIVES

The diversity of GABAergic neurons and neural communication elements

Z. Josh Huang and Anirban Paul

Abstract | The phenotypic diversity of cortical GABAergic neurons is probably necessary for their functional versatility in shaping the spatiotemporal dynamics of neural circuit operations underlying cognition. Deciphering the logic of this diversity requires comprehensive analysis of multi-modal cell features and a framework of neuronal identity that reflects biological mechanisms and principles. Recent high-throughput single-cell analyses have generated unprecedented data sets characterizing the transcriptomes, morphology and electrophysiology of interneurons. We posit that cardinal interneuron types can be defined by their synaptic communication properties, which are encoded in key transcriptional signatures. This conceptual framework integrates multi-modal cell features, captures neuronal input–output properties fundamental to circuit operation and may advance understanding of the appropriate granularity of neuron types, towards a biologically grounded and operationally useful interneuron taxonomy.

Understanding the biological basis of neuronal diversity is necessary for deciphering neural circuit organization and function[1,2]. In the mammalian cerebral cortex, GABAergic interneurons regulate the balance and dynamic organization of pyramidal neuron ensembles[3], which in turn mediate myriad information processing streams and output channels[4,5]. The spectacular diversity of GABAergic neurons enables an elaborate division of labour in deploying a rich repertoire of inhibitory control mechanisms to shape highly nuanced spatiotemporal dynamics of cortical circuit computation[6,7]. This diversity of cortical interneurons has been apparent since Ramon y Cajal, as the morphology of these 'short axon cells' were more readily visualized with the Golgi stain[8]. Seminal discoveries have revealed the stunning specificity of interneurons in terms of their synaptic connectivity (for example, their selectivity in synaptic partners and subcellular compartments), their physiological characteristics (such as intrinsic and especially synaptic properties) and functional properties in circuit operations (such as their temporal integration of inputs and contributions to

network oscillations)[3,6,7,9]. These studies have also revealed immense, and often seemingly intractable, phenotypic variations along multiple axes that defy any simple classification scheme[10]. Consequently, the issue of interneuron diversity remains contentious[11]. There have been continued debates on how interneuron 'types' should be defined and classified and the scope of their diversity (that is, how many types there are and with what granularity) — debates that are relevant to neuron types in general across the CNS. Indeed, many important problems in neuroscience can be attributed to the ambiguity of the definition of 'neuron types' in the brain circuits under investigation[12]. At the core of these debates is the issue of whether neuronal identity and classification can be grounded on underlying biological mechanisms and principles or are destined to remain arbitrary and operational.

Given that the multi-modal and multi-dimensional phenotypes of nerve cells include not only morphology, connectivity patterns, physiological properties and gene expression profiles but also developmental history and ultimately circuit and behavioural function, the gaps in our current knowledge and

understanding are not surprising. Despite major advances in past decades, until recently investigators remained technically underpowered to measure multi-modal cell features. For example, sparse and partial (that is, non-systematic and incomplete) reconstructions of individual cells preclude truly quantitative analysis of cell morphology. Most electrophysiological recordings sample limited and often biased cell populations and only a fraction of the physiological parameters among the diverse input–output (I/O) transformation properties. Serendipitously identified molecular markers represent a minute fraction of gene expression profiles. Thus, overall, these studies remain severely limited in resolution, robustness (which includes reproducibility across investigators) and comprehensiveness.

The conceptual obstacle to understanding neuronal diversity manifests at multiple levels. First, neurons display substantial multi-dimensional variations, of which some are discrete whereas others seem continuous[13,14]. As we often cannot distinguish between biologically meaningful variations and stochastic or technical variations, it is difficult and often arbitrary to set boundaries for cell clustering by adjusting statistical and algorithmic parameters. Second, the inherently multi-modal nature of cell phenotypes requires integrative classification across modalities, but it is not clear whether different cell features, as defined and measured by current approaches, indeed co-cluster; this raises the issue of whether congruent multi-modal classification is ultimately achievable. Third, it seems intuitive that neurons should be viewed as members of a type if they serve a common function, and cell typing is useful only if it ultimately helps understand such function or functions. In practice, however, cell function per se is often not readily definable and emerges only at the circuit level, which is difficult to study and is often indirectly linked to behavioural performance. At the core of these challenges is the lack of an overarching framework of neuron-type identity that integrates and explains multi-modal variations relevant to circuit operation and that is grounded on biological mechanisms and principles.

In the past few years, spectacular advances in single-cell analysis have finally crossed

several technical thresholds and begun to generate high-resolution, quantitative and comprehensive data sets on single-neuron transcriptomes, morphology and electrophysiology. Innovative statistical and computational analyses now enable clustering and typing along each modality and drive efforts to integrate across modalities. In this Opinion article, we review large-scale single-cell RNA sequencing (scRNAseq) studies that have unveiled a working draft of a comprehensive transcriptomic interneuron taxonomy and the related single-cell morphology and electrophysiology data sets that contribute to multi-modal measurements. We discuss the current limitations in achieving correspondence among multi-feature descriptions and difficulties in cross-modality cell clustering, which may reflect the lack of a conceptual framework that distils the overarching properties that encapsulate and extend beyond traditionally described phenotypes. We highlight a recent finding that defines cardinal interneuron types as canonical neural 'communication elements' with characteristic I/O transformation properties that are encoded by transcriptional signatures of key gene families. We elaborate on the implications of this emerging framework of neuronal identity in integrating multi-modal phenotypes and in facilitating cell-type classification based on biological principles beyond operational management. We discuss outstanding challenges and opportunities for future progress.

Classification and taxonomy

Single-cell transcriptomics. To a considerable extent, the phenotypes and properties of cells derive from their patterns of gene expression. Transcription profiles of neuronal subpopulations have been characterized through microarray[15] or RNAseq[16], but studies using bulk cell populations are inherently limited in resolving heterogeneity. Pioneering studies of single-cell expression started more than two decades ago[17], and the later invention of scRNAseq[18] was soon applied to nerve cells[19]. However, it is only in the past 4 years that large-scale scRNAseq technology has been able to provide a comprehensive molecular analysis of neuronal diversity[20,21].

A pilot study of ~1,600 cells from the mouse primary visual cortex used scRNAseq and statistical clustering to identify 49 cell types, including 23 GABAergic transcriptomic types (t-types)[22]. A subsequent large-scale study of ~23,800 cells incorporating multiple transgenic lines targeting neuronal subpopulations suggested 113 cortical cell types; these include ~60 GABAergic types conserved between visual and frontal cortex and 56 glutamatergic types that are mostly distinct between these two areas[23] (FIG. 1). This landmark data set represents the most comprehensive survey of cortical transcriptional types, establishing a scaffold for transcriptome-based classification and a working draft of cortical cell taxonomy, including GABAergic interneuron taxonomy. The major subclasses, types and their hierarchical relationships are largely consistent with previous studies[7,24,25]. Some of the finer types and branches (for example, *Sst–Chodl*, *Sst–Calb2–Pdlim5* and *Sst–Chrna2–Glra3*) are also consistent with previous studies[26–28] (FIG. 1). Certain clusters may correspond to newly discovered types; for example, one of the four clusters defined by *Vip* plus *Chat* expression might correspond to a type of interneuron that directly excites layer 1 interneurons and layer 2/3 pyramidal neurons (as described in a preprint report)[29].

However, the taxonomy of cortical GABAergic neurons is not settled. The ~60 'atomic' GABAergic t-types at the base of the hierarchy[23] are probably fluidic and may well be modified by future scRNAseq and orthogonal data sets. Currently, the choices of statistical algorithms and the stringency of cluster definitions remain largely arbitrary[30]. For example, the number of *Sst* transcriptional clusters can vary by a factor of 2 depending on the stringency of clustering (30 when less stringent and 15 when more stringent)[23]. These results highlight the issue of the extent to which 'neuron types' are statistical or biological.

Moreover, several results from the large analysis seem inconsistent with previous studies. For example, decades of developmental studies have established a clear division of transcription programmes between the interneurons derived from the medial and caudal ganglionic eminence (MGE and CGE, respectively)[9,31–33]. In particular, a transcription cascade involving *Nkx2.1* and *Lhx6* drives the MGE clade but not the CGE clade of interneurons[31,34,35]. Unexpectedly, a group of *Lamp5* subclass interneurons with prominent *Lhx6* and *Nkx2.1* expression were classified as CGE-derived putative neurogliaform cells (NGFCs)[23]. However, this *Lamp5–Lhx6* type was found to be transcriptionally similar to L5/6 chandelier cells (CHCs) known as the CHC2 type that were fate-mapped from *Nkx2.1*-expressing progenitors in the MGE clade in another study[28]. It will be important to resolve this discrepancy,

as the distinction between NGFC and CHC types is positioned towards the top of the hierarchy of interneuron classification. One possibility is that CHC2s[28] are in fact NGFCs and that *Nkx2.1* and *Lhx6* are not strictly restricted to the MGE lineage and are also expressed by certain CGE cell types; this would cast a doubt on the stringency of the core transcription programmes that separate the fundamental MGE and CGE lineages. Another possibility is that *Lamp5–Lhx6* cells are in fact MGE-derived deep-layer CHCs. In this scenario, non-essential transcriptional features (such as those derived from technical or tangential factors) could have overtaken key biologically meaningful signals and misled the clustering algorithm. It is also possible that CHC2s may represent or contain an MGE-derived but uncharacterized cell type. Resolving this discrepancy will not only clarify the distinction between two well-characterized bona fide types (NGFCs and CHCs) but also may reveal the power of, and glitches in, statistical clustering. For example, clustering algorithms may need to be supervised so that certain foundational core transcripts (such as *Nkx2.1* and *Lhx6*) have greater 'voting power' (or are more highly weighted) in clustering of cell identities and relationships than other genes.

The current fluidic nature of transcriptomic clustering is at least in part due to our incomplete understanding of the biological basis of neuron-type identity and granularity and the paucity of 'ground truth' knowledge. Such understanding probably requires the integration of multi-modal cell features, recognition of biologically meaningful variations in each feature and the elucidation of their mechanistic bases.

Single-cell morphology. Describing neuron types on the basis of their morphology seems intuitive. However, the vast diversity and seemingly endless variations of neuronal shapes present major challenges in morphological tracing and analysis. A century after Ramon y Cajal described 'short axon cells', rigorous, quantitative and scalable analysis of interneuron morphology remains difficult[11].

Quantitative single-neuron anatomy requires overcoming four technical hurdles. The first is labelling — to systematically, reliably and completely label specific sets of individual neurons. The second is imaging — to achieve large-volume imaging with axon resolution (~100 nm) across the brain. The third is reconstruction — to convert large image stacks into digital data sets of single-neuron morphology. The fourth
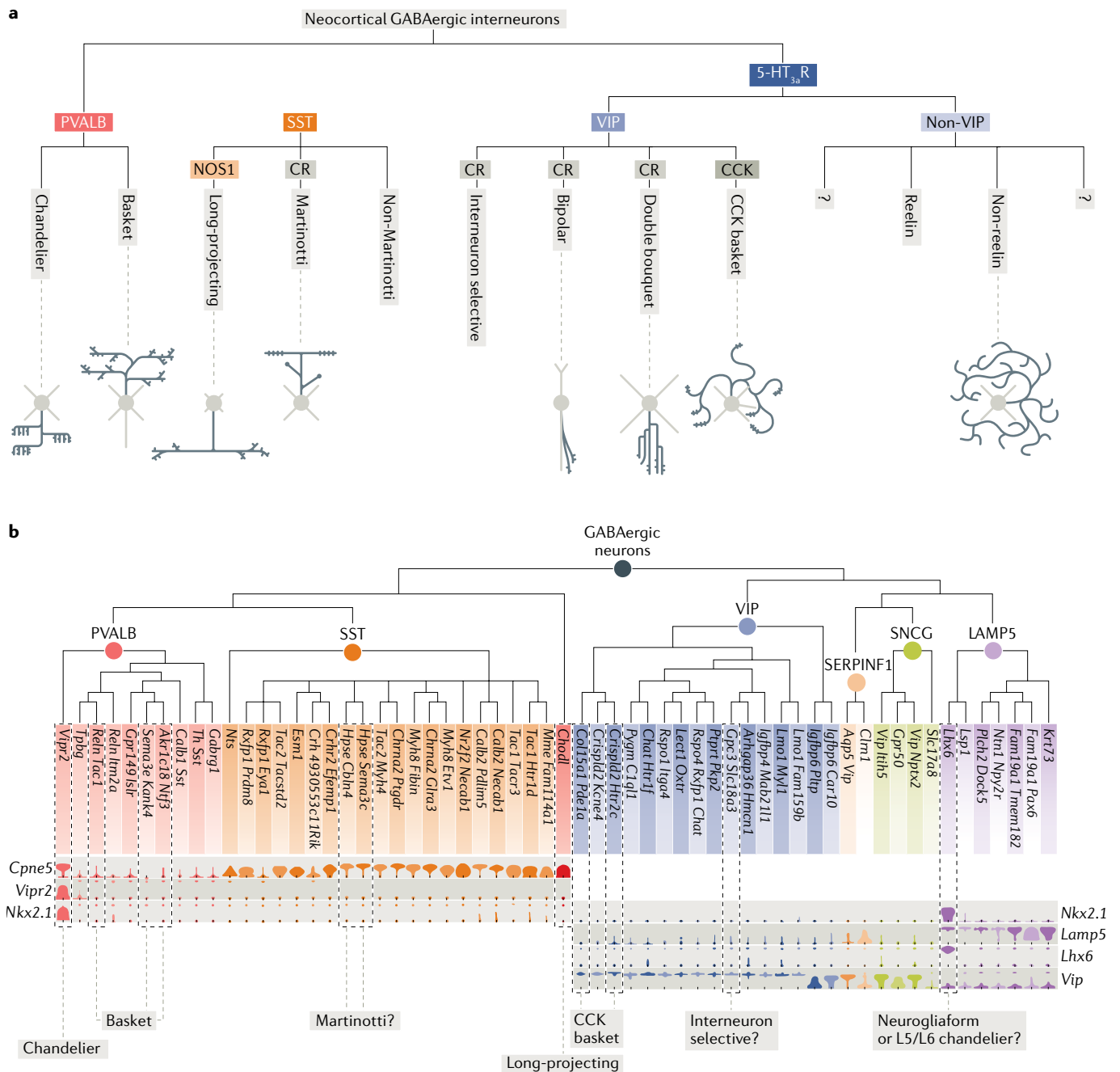
Fig. 1 | **A work draft taxonomy of transcriptomic neuron types of the cortical GABAergic system. a** | Major GABAergic subclasses and cell types recognized by classic anatomical, physiological, molecular and developmental studies. Schematics of the characteristic morphology of these cell types are also shown. Light grey lines represent dendrites; dark grey lines represent axons. **b** | Current taxonomy of transcriptomic types in the primary visual and anterolateral motor cortex. Note that the major types and branches are consistent with classic studies. Currently only a few of the 60 'atomic types' are well correlated with the anatomically and physiologically defined types. '?' denotes that the correlation between the transcriptomic type and the indicated morphological type is not yet firmly established. A major discrepancy is the *Lamp5–Lhx6*-type cells, which were clustered as caudal-ganglionic-eminence-derived neurogliaform cells[23] but were recognized as medial-ganglionic-eminence-derived chandelier cells (CHCs)[28]. 5-HT$_{3A}$R, serotonin 3A receptor; CCK, cholecystokinin; CR, calretinin; NOS1, nitric oxide 1; PVALB, parvalbumin; SST, somatostatin; VIP, vasoactive intestinal peptide. Part **b** is adapted from REF.[23], Springer Nature Limited.

is analysis — to register neuronal morphology with an appropriate spatial coordinate framework and to extract, quantify and classify biologically relevant attributes (for example, those related to neural connectivity). Most studies in past decades using traditional methods can recover only partial interneuron

morphologies that are not well registered to a common spatial coordinate framework. These limitations make it difficult to achieve quantitative analysis, to distinguish random or technical variation versus reliable and meaningful features, and to compare results from different investigators.

One large-scale study in juvenile rat brain slices digitally reconstructed 1,009 cortical neurons[24]. Statistical analysis combined with literature mining and expert annotation classified these into 55 morphological types (m-types), including ~40 interneuron types with their associated laminar

locations. In another study using brain slices of adult mouse visual cortex from a large set of transgenic lines, 254 spiny and 207 aspiny neurons were reconstructed[36]. Using unsupervised hierarchical clustering, 19 spiny m-types and 19 aspiny m-types were identified. Together, these efforts represent the most comprehensive morpho-physiological analysis of cortical interneurons to date. Despite this progress, brain-slice preparations present inherent limitations for complete morphological reconstruction (for example, many of the spiny neuron axons are transected and cannot be visualized), precise spatial registration and truly quantitative analysis.

The recent integration of genetic labelling and axon-resolution, large-volume imaging has begun to overcome major technical hurdles of single-neuron anatomy in rodent brains. In one platform, a high-speed two-photon microscope is integrated with sparse viral labelling of neurons and computational tools for large-scale image analysis[37]. In another platform, dual-colour fluorescence micro-optical sectioning tomography (dfMOST) allows axon-resolution imaging and cell-resolution spatial registration of individual genetically labelled neurons[38]. A recent fMOST (fluorescence micro-optical sectioning tomography) study achieved complete reconstruction of over 60 CHCs in the mouse neocortex. Analyses of their laminar position and the distribution of their dendritic and axonal arborizations suggested that multiple CHC subtypes can be distinguished by different patterns of I/O connectivity[39]. These technical advances signal the rise of high-resolution, quantitative and scalable single-neuron anatomy in the mouse brain.

A major current bottleneck is morphological reconstruction, which is mostly achieved by manual procedures that are particularly labour intensive. Innovations in more automated reconstruction and registration will be necessary to generate high-throughput and comprehensive data sets[40,41]. In terms of analysis, conceptual as well as technical questions remain. How can or should biologically meaningful features be distinguished from stochastic variations? To what extent can morphology inform cell typing? As morphology is a proxy for connectivity, how should it be interpreted in the context of connectivity?[42]

Electrophysiology. The electrophysiological properties of neurons are proximal to their roles in circuit operation but are even more difficult to measure comprehensively and quantitatively at scale, especially in vivo. Different neuron types probably mediate distinct sets of I/O transformations that are supported by their intrinsic, synaptic and network properties, which span orders of spatiotemporal scales (from dendritic spines to axon terminals, and from submilliseconds to seconds) and are further influenced by brain state and behaviour.

Largely owing to technical limitations, only a small set of physiological properties can be routinely measured, the most common being intrinsic properties at the cell soma and a limited set of synaptic properties, predominantly in brain slices in vitro. For example, analysis of the responses of >3,900 rat cortical neurons to a standardized battery of stimulation protocols revealed 11 electrophysiological types (e-types), including 10 inhibitory types and 1 excitatory type[24]. Using a similar approach, another impressive study of 1,938 neurons in the visual cortex of a set of transgenic mouse lines (in which neuronal subpopulations were labelled) classified 4 excitatory e-types and 13 inhibitory e-types[36]. These data sets are highly valuable, although the intrinsic responses measured at the cell soma in vitro provide a rather limited glimpse of the sophisticated spatiotemporal operations of neurons in their network niche.

Measuring synaptic properties requires multiple simultaneous patch-clamp recordings, which are technically demanding and even more difficult to scale. An impressive profiling of cortical cell types using octuplet whole-cell recording in adult mouse cortex mapped connectivity between more than 11,000 pairs of identified neurons[43] and identified 15 types of interneurons, each exhibiting a characteristic pattern of connectivity with other interneuron types and pyramidal neurons. However, as each neuron receives inputs from and transmits outputs to multiple presynaptic and postsynaptic neurons, respectively, a more comprehensive profiling of synaptic signalling is prohibitively challenging, even with octuplet patch recording by expert physiologists.

The firing patterns of hippocampal interneuron types during various in vivo network oscillations are highly characteristic[44,45], but because network oscillations are less easily discernible in the neocortex, similar measurements are not yet feasible for cortical interneurons. Therefore, current measurements of electrophysiological properties provide only a limited view of the sophisticated and probably characteristic spatiotemporal and network operations. As most common data sets on intrinsic properties in vitro may not reflect the essence of neuronal physiological operations, it is perhaps not surprising that they have not been particularly informative in neuronal classification. Proper distinction of physiological neuron types may require the innovation and integration of multiple novel technologies to distil overarching functionally relevant features — for example, by measuring the I/O properties in the context of circuit operation.

Multi-modal data correspondence
With the broad application of scRNAseq, it is now possible to begin to correlate transcriptomic, morphological and physiological data sets with the aim of building a multi-modal cell-classification scheme. One such integrative approach is Patch-seq, which combines whole-cell patch-clamp recordings, scRNAseq and morphological characterization[46,47].

Another approach is to integrate multi-modal measurements of the same neuronal subpopulation labelled by a common mouse driver line. The same large-scale analysis of adult mouse visual cortex[36] used 28 Cre driver lines to enable broad coverage, selective targeting of rare populations and linking to previous single-cell transcriptomic data sets[23] — the intrinsic properties of >1,900 neurons (928 spiny and 1,010 sparsely spiny) were analysed and the morphologies of 461 neurons (254 spiny and 207 aspiny) were partially reconstructed. This analysis suggested a trend of correspondence between the identified 17 e-types, 38 m-types and 46 jointly classified me-types; for example, thick tufted cells in L5 (called the spiny7 m-type) corresponded quite well to an e-type called Exc_3. However, in many cases, the correspondence between e-types and m-types was not strong. To examine the relationship between t-types to e-types and m-types, the same driver lines that each corresponded to a small number of t-types were used to characterize e-types and m-types[36]. The best example of multi-modal correspondence is the L5/6 *Sst–Chodl* long projection GABA neurons, which are captured by *Nos1–Sst* intersection[27,28]. In general, however, the correspondence among t-types, e-types and m-types is relatively coarse. Thus, the extent to which the 60 t-types[23] correspond to the 17 e-types and 38 m-types and 46 me-types[36] is currently not well established.

Given these results, one might conclude that there is probably not a tight correspondence among the transcriptomic, morphological and electrophysiological features of GABAergic neurons. However, such a conclusion is premature for several

reasons. Given that intrinsic properties measured using arbitrary stimulation parameters in vitro poorly represent the overall electrophysiological features and the limitations of morphological analysis of partially reconstructed neurons from brain-slice preparations, the current relatively weak correspondence between e-types and m-types is not surprising. Furthermore, as morphological and physiological measurements have far lower resolutions than do transcriptomic measurements (which, given the number of mRNA transcripts in each cell, have thousands of dimensions), the relatively
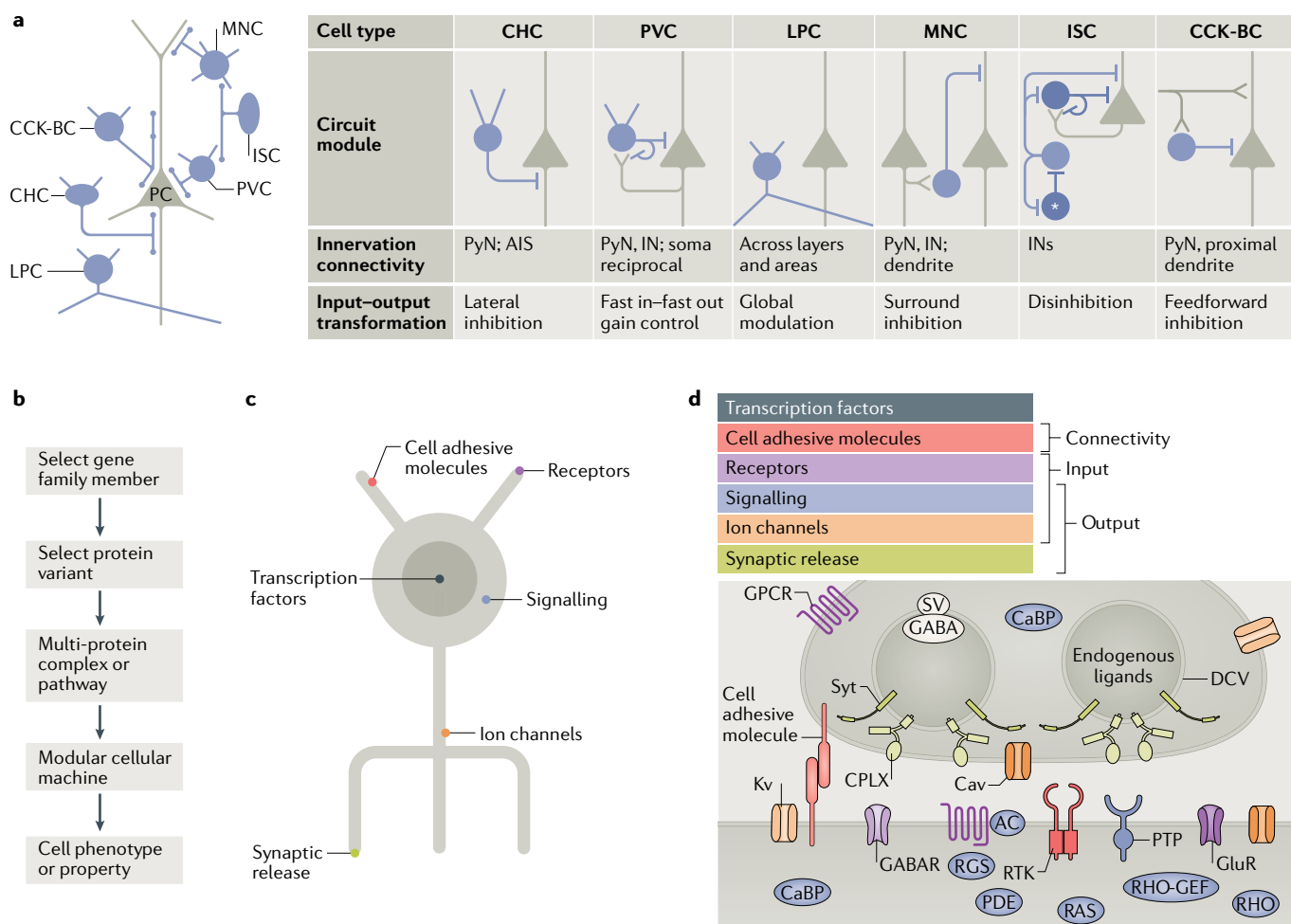
weak correspondence between the t-types, e-types and m-types is also not unexpected. The current ambiguity and disparity in cell clustering across modalities might reflect the lack of a conceptual framework that captures the essence of neuron type and distils the overarching properties that encapsulate or extend beyond the traditionally described cell phenotypes.

Communication elements
Transcription profiles of cellular machines shape I/O properties. Despite progress in large-scale single-cell transcriptomics, the problem of neuronal diversity and

classification is unlikely to be solved without understanding the biological principles that underlie neuronal identity and the organization of cell types. Considerable progress has been made by analysing high-resolution single-cell transcriptomes through leveraging the deep, although incomplete, knowledge in GABAergic neuron anatomy and physiology.

One study[28] differed from other single-cell transcriptomic studies in three important ways: experimental design, RNAseq method and data analysis (FIG. 2). Unlike efforts to discover diversity and classify neurons using unsupervised

a

Cell type	CHC	PVC	LPC	MNC	ISC	CCK-BC
Circuit module						
Innervation connectivity	PyN; AIS	PyN, IN; soma reciprocal	Across layers and areas	PyN, IN; dendrite	INs	PyN, proximal dendrite
Input–output transformation	Lateral inhibition	Fast in–fast out gain control	Global modulation	Surround inhibition	Disinhibition	Feedforward inhibition

b

Select gene family member → Select protein variant → Multi-protein complex or pathway → Modular cellular machine → Cell phenotype or property

c

Cell adhesive molecules · Receptors · Transcription factors · Signalling · Ion channels · Synaptic release

d

Transcription factors	
Cell adhesive molecules	Connectivity
Receptors	Input
Signalling	
Ion channels	Output
Synaptic release	

Fig. 2 | **Transcriptional signatures of synaptic communication delineate cardinal GABAergic neuron types. a** | Six cardinal types — long projection cells (LPCs), chandelier cells (CHCs), PV-basket cells (PVCs), cholecystokinin-expressing basket cells (CCK-BCs), interneuron selective cells (ISCs) and Martinotti cells (MNCs) — are distinguished by their characteristic innervation of cellular and subcellular targets. Where data are available, these cardinal types manifest distinct input–output connectivity patterns and further display distinct intrinsic, synaptic and network properties indicative of mediating specific forms of input–output transformation. **b** | Cell phenotypes and properties emerge from the workings of macromolecular machines (cellular modules) that consist of multi-protein complexes. As molecular components of these complexes, protein variants with different biochemical and biophysical properties encoded by members of a given gene family can customize

cellular modules and cell-type-specific properties. **c** | Examples of cellular modules that shape neuronal connectivity, synaptic transmission, electrical signalling, intracellular signal transduction and gene transcription. **d** | Six categories of gene families shape a set of membrane-proximal molecular machines that customize the input–output connectivity and transformation properties of different GABAergic cardinal types. AC, adenylyl cyclase; AIS, axon initial segment; CaBP, calcium-binding protein; Cav, voltage-gated calcium channel; DCV, dense core vesicle; GluR, glutamate receptor; GPCR, G protein-coupled receptor; IN, interneuron; Kv, voltage-gated potassium channel; PDE, phosphodiesterase; PTP, protein tyrosine phosphatase; PyN, pyramidal neuron; RGS, regulator of G protein signalling; RHO-GEF, RHO guanine nucleotide exchange factor; RTK, receptor tyrosine kinase; SV, synaptic vesicle; Syt, synaptotagmin. Adapted with permission from REF.[28], Elsevier.
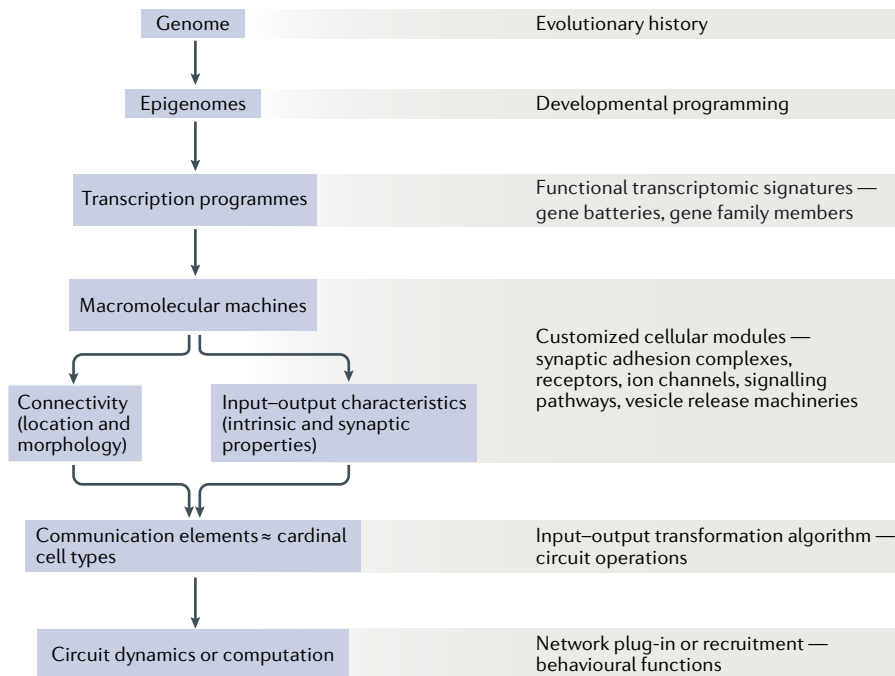
Fig. 3 | **A conceptual framework of neuronal identity.** A neuron type can be defined as a canonical neural communication element that mediates characteristic input–output transformations and that is encoded by key transcription signatures that customize a congruent set of cellular machines. This scheme integrates the anatomical, physiological, functional and developmental genetic features that together define neuronal cell types. It emphasizes cardinal neuron types that are reliably generated in each member of a species, build species-stereotyped circuit templates and are likely rooted in the genome. Not shown is the notion that these cardinal types are probably further diversified and customized (for example, by neuronal activity and experience) to shape circuit elements characteristic of individual organisms. Features in parentheses are commonly measured cell properties but are reflections of core features of cell types. See text for detailed description.

statistical clustering of transcriptomes, Paul et al. focused on understanding the molecular basis of neuronal identity by analysing the transcriptomes of six cardinal, phenotypically characterized populations (PCPs) of GABAergic neurons captured by intersectional genetic targeting. These PCPs were described according to previous anatomical, physiological and developmental research, providing a solid basis for the interpretation of and correspondence to the transcriptomic data set. This study also used a modified version of CEL-seq (cell expression by linear amplification and sequencing), called DIVA-Seq (double in vitro transcription with absolute molecule counts sequencing)[48], to detect ~9,000 genes expressed per single cell and to count mRNA transcripts in single cells. This transcription profiling method achieves substantially higher resolution than Drop-Seq and 10x Genomics platforms and more accurate mRNA-level measurements than Smart-Seq. The study aimed to extract key transcription features that contribute to, explain and predict cell properties. This strategy was based on the premise that most cellular properties arise from the operations of macromolecular machineries (or 'cellular modules') that each consists of multiple interacting protein components[49]; an individual protein component is often implemented as a member of a gene family that contains several variants, each differing in biochemical and biophysical properties that modify their shared function. Thus, the differential expression of certain variants of protein components may tailor the features of these macromolecular machineries, thereby customizing the phenotypes that together characterize a cell type (FIG. 2).

Paul et al.[28] applied a supervised machine learning-based algorithm (MetaNeighbour[50]) to screen the ~620 HGNC (Human Genome Nomenclature Committee) gene families for those whose differential single-cell expression reliably distinguish these PCPs. Strikingly, PCPs were best distinguished by the expression of members of a set of only ~40 gene families. These gene families comprise six functional categories that include cell-adhesion molecules, neurotransmitter and modulator receptors, ion channels, regulatory components of membrane-proximal signalling pathways, neuropeptides and vesicular release components, and transcription factors. The expression of certain combinations of gene family members from across these functional categories seems to shape the characteristics of a molecular scaffold at the cell membrane that contribute to presynaptic and postsynaptic properties — that is, I/O characteristics (FIG. 2). Using the expression of these ~40 gene families to discriminate interneuron type in data sets from two other studies with disparate experimental designs[21,22] was similarly effective. These findings suggest that cardinal neuron types can be defined as neural communication elements that are rooted in transcriptional programmes that orchestrate functionally congruent expression across certain gene families to customize the patterns and properties of I/O transformation. This is consistent with an independent proposal to define projection neuron types on the basis of I/O relationships[51].

An overarching and mechanistic framework of neuronal identity. The concept of 'modular organization' of cortical architecture dates back more than four decades[52–54]. Yet the neurobiological basis of 'canonical circuits' has remained elusive and contentious[55], largely owing to the lack of quantitative, comprehensive and multi-modal data and a limited understanding of neuron types. The definition of neuron type as a neural communication element rooted in a transcription programme is a higher-level abstraction that integrates multi-modal cell phenotypes, provides a framework for understanding neuronal diversity and may facilitate biologically principled cell-type classification beyond statistical algorithms and arbitrary criteria (FIG. 3). The two fundamental features of a communication element are its connectivity (that is, communication partners) and I/O transformation (that is, communication mode), which encapsulate and extend the more conventional descriptions of cell features.

Although location and morphology have been prime descriptors of neurons, they reflect the more fundamental property of connectivity. Thus, seemingly intractable variability in morphology may belie the more constant patterns of connectivity between presynaptic and postsynaptic elements[56]. Indeed, computational algorithms have been developed to reliably identify m-types from dense connectome data sets[57,58]. In terms of physiological features, the intrinsic, synaptic and network properties of a neuron type probably all serve in the transformation of information

from various synaptic input patterns to appropriate synaptic output patterns, often characterized by a cell-intrinsic style of neurochemical release. Defining neuronal types on the basis of their 'function' seems intuitive and the most relevant to nervous system operation[9,51], but neuronal functions manifest at multiple levels that carry different meanings, vary according to behavioural context, and thus seem not to be well suited for use in defining cell types. A key feature of the communication element concept is that it seamlessly incorporates the functional attributes of neuron types across levels. Neuronal communication elements with specific I/O transformation algorithms carry out a specific set of definable circuit operations (such as feedforward inhibition, lateral inhibition, disinhibition, gain control or γ-oscillations) — their proximal level functions. These elements can be task-dependently recruited into larger brain networks for systems-level information processing and computation, representing their distal behavioural and cognitive functions[59,60]. Thus, the communication element concept incorporates the functional definition of neuron types by clarifying and integrating different levels of functional explanations. The cardinal identity of a cell is probably already instilled at the postmitotic stage during development[9,61–64], and its subsequent phenotypic differentiation and expression are regulated by extrinsic signals such as cell interactions and neural activity. Importantly, key communication properties are shaped by specialized

cellular modules that are encoded through transcriptional programmes. Although the transcription profile of a cell is influenced by extrinsic factors of the cellular milieu (including neural activity)[65], core transcription programmes are primarily shaped by cell-intrinsic epigenomic landscapes (BOX 1), which mainly derive from differential developmental programming of the genome. Therefore, the communication element definition may coherently integrate molecular, anatomical, physiological, functional and developmental genetic features (FIG. 3); it is a testable hypothesis (BOX 2).

An intriguing corollary is that although communication is inherently a cell-extrinsic feature that is directly affected by communication partners, it is rooted in cell-intrinsic transcription programmes. Therefore, as part of its core identity, a cardinal cell type may be driven by its genetic information to establish characteristic connectivity patterns with other appropriate cell types[63]. Cell-intrinsic transcription programmes (derived largely from epigenomes) might therefore represent an intermediate step through which the genome ultimately orchestrates the self-assembly of canonical features of neural circuits. Thus, the communication element definition of neuron types should facilitate the delineation of higher-level modular organization of neural circuit architecture[24,53] by clarifying the developmental genetic basis of elemental building blocks.

As such, the recent finding that there has been relatively poor correlation between

currently described m-types and e-types is perhaps not surprising, as morphology indirectly reflects connectivity, whereas electrophysiological properties measured at somata are just a few of many contributing factors that determine I/O transformations. Key connectivity and I/O features might potentially be highly correlated once they can be measured at a high enough resolution. Certain morphological features might be more tightly linked to connectivity (for example, CHC axon arborizations[39]), and certain intrinsic and synaptic features might be better linked to I/O properties (for instance, fast-spiking basket cells show fast in–fast out signal dynamics[66]). Once identified, tighter correlations in specific subpopulations may emerge.

Granularity of neuron types
In addition to understanding the nature of neuronal identity, the appropriate granularity of cell types needs to be defined for a meaningful classification and taxonomy[14]. The question of how many cell types a given brain region contains is frequently asked but will not have a clear answer until investigators can determine how finely and firmly cell types should be distinguished to be useful for understanding circuit function. Inherently related to the issue of neuron-type identity, the question of granularity presents similar technical and conceptual challenges at a finer resolution.

Even neurons recognized as members of a given subclass or type often manifest substantial variation (discrete and continuous) along multiple modalities. Whether and when investigators should continue to finely cluster cells along to each modality to define more 'subtypes' is unclear. For example, single-cell transcriptome data sets manifest discrete as well as continuous variations[13,22]. A major question is which variations are biologically meaningful and which are not (for example, they may instead be technical or stochastic). Similarly, morphological variations between neurons of the same class may seem continuous and intractable[11] but may belie the more specific and discrete connectivity patterns between certain presynaptic and postsynaptic elements[56,67]. In addition, the importance of the variation of intrinsic physiological properties may be interpretable only in the context of their effect on I/O transformation. Although some argue for an emphasis on discrete variables that shape a hierarchical taxonomic tree with binary splits[14], others suggest continua as a necessary and probabilistic description of cell diversity[13,68].

Box 1 | Single-cell epigenomics

The transcriptome of an individual mature neuron is the transcriptional output of its epigenome, which consists of a record of the chemical modifications of the genome (including DNA methylation and histone modification and so on) that influence how genes are expressed in that neuron. Epigenomes in the organism are customized from a singular genome, largely through developmental genetic programming. In addition to single-cell RNA sequencing (scRNAseq), technical advances now enable the delineation of genome-wide DNA-methylation patterns[83] and chromatin accessibility (using the assay for transposase-accessible chromatin (ATAC) sequencing[84]) in single cells. Assaying the DNA-methylation pattern from ~6,000 single mouse and human frontal cortex nuclei, cell-type-specific regulatory elements with differential methylation patterns were identified across neuron types[83]. In another complementary study, potential master transcriptional regulators were inferred from chromatin accessibility assay in ~15,000 single nuclei using ATAC-seq[84].

Although these single-cell epigenomics techniques have yet to reach the resolution and scale of scRNAseq, they have begun to resolve major cortical cell classes and subclasses (including glutamatergic neurons, GABAergic neurons, parvalbumin-expressing neurons, somatostatin-expressing neurons, glial cells and so on). With further technical improvement, integrated analyses of single-cell transcriptomic and epigenetic data sets[62] will not only facilitate molecular cell typing but also may provide deeper insight into the underlying gene-regulatory basis for cell typing. For example, developmental time series data sets may reveal the emergence of epigenomic landscapes that establish an essential set of transcriptional enhancers, which orchestrate core transcriptional programmes to shape key gene family expression profiles, thereby customizing the synaptic communication properties that define cell types (FIG. 3).

The communication element definition of neuron type should be viewed as a testable hypothesis, with the prediction that input–output (I/O) transformation properties correlate with key transcriptional signatures and together define cardinal neuron types.

Recently developed tools that combine genetic targeting[27,85], anterograde, retrograde and trans-synaptic viral tracing (such as TRIO; tracing the relationship between input and output[86]), and optogenetic mapping are well poised to facilitate the discovery of the I/O streams of neuronal subpopulations[2,51]. In parallel, dense connectomics data sets obtained using volume electron microscopy (EM) may reveal the complete connectivity of individual neurons and enable investigators to discern key patterns that are characteristic of cardinal types. Current efforts on EM reconstruction of a cortical column may reveal the local connectomes of interneurons and begin to allow investigators to examine the relationship between connectivity and cell identity[87]. Larger-volume EM connectomes will be necessary to fully examine the I/O streams of cortical neurons.

There is currently no single method to measure I/O transformation. Instead, different components, such as synaptic inputs and integration, spike initiation and propagation, and synaptic release, are measured separately only in a few abundant and accessible cell populations (for example, parvalbumin-expressing fast-spiking basket cells[66]). Alternatively, I/O transformations can be probed on the basis of the position of neurons in cellular-resolution-defined anatomical pathways via advances in subpopulation-targeted and wiring-dependent optogenetic control, activity readout and perturbations[51]. However, despite much progress, large-scale measurement of I/O transformations in neuronal subpopulations in vivo remains a considerable challenge and may require entirely new technologies.

Importantly, I/O transformation properties seem to be encrypted in key transcription signatures[28]. Proper analysis of high-resolution transcriptomes may guide the characterization and validation of neuronal I/O patterns and facilitate the discovery of neuron types (that is, beyond statistically defined transcriptomic types). Recent advances in spatial transcriptomics and its increasing integration with modern anatomical and imaging tools[88–92] may facilitate the correlation of I/O properties with transcriptional signatures. Applying this strategy to a modest set of cardinal types, such as glutamatergic pyramidal neurons, seems feasible. Once a correspondence between key transcription signatures and I/O properties has been established, scalable high-resolution transcriptome analysis may guide the large-scale discovery of cardinal neuron types.

These debates uncover a deeper conceptual issue: what level of granularity of cell-type definition is necessary for understanding circuit function?

The communication element concept of neuron types may provide a mechanistic and functional basis for defining the appropriate granularity of a neuron type taxonomy. A better understanding of the I/O transformation properties of a broader neuron type may be necessary to define finer granularity. In some cases, seemingly subtle variations may endow novel properties of a generic type of I/O transformation to a subset of cells, thereby warranting their classification as a separate subtype (for example, the 15 types of retinal bipolar cell[69]). In other cases, a continuous I/O transformation that determines the circuit operation of a communication node may be achieved by implementing certain continuously variable features to its constituent cell members (for example, orientation-selective neurons in primary visual cortex all encode orientations but each has a particular orientation preference). It is possible that certain cardinal types might have inherent tendencies and mechanisms to manifest further variations in multiple modalities and thus to diversify into finer varieties that together better serve their overarching communication function. A satisfying cell-type taxonomy might need to integrate hierarchical trees of discrete cell classes and types with probabilistic and continuous features along certain branches. Proper recognition of cell-type granularity will also help to distinguish variations that are related to cell identity from those attributable to different physiological or pathological states.

Conclusions and perspectives

The importance of discovering and understanding neuronal diversity is now well recognized. We suggest that uncovering the biological basis of neuron-type identity and granularity is necessary to decipher neuronal diversity and to achieve a satisfying classification and taxonomy. High-throughput scRNAseq and epigenomic analyses will accelerate the generation of massive data sets, laying out the molecular landscape for systematic cell-type discovery and classification. The proposed framework here — that each cardinal neuron type can be defined as a neural communication element with characteristic I/O transformation properties that are encoded by key transcriptional signatures — posits that neuron types are not just statistical constructs shaped by arbitrary criteria but are inherently biological entities built with molecular genetic mechanisms that can be understood through fundamental principles. High-throughput single-cell analyses grounded on the proposed neuron identity framework should substantially clarify the working draft of interneuron taxonomy and drive cell-type discovery across the cerebral cortex, a strategy with broad implications for exploring cell diversity in the CNS.

The two basic features of neuronal communication, connectivity and I/O transformation, cannot currently be measured at scale (BOX 2). Technical improvements in I/O tracing and dense connectomics will facilitate the scalable measurement of neural connectivity. In parallel, continued innovation in genetic targeting, optogenetic manipulation, activity readout and perturbation will enable the characterization of I/O transformation in greater numbers of neuronal subpopulations. Fortunately, it is increasingly evident that transcriptomes and epigenomes contain highly rich information regarding cell identity and physiological state[70]. Indeed, the transcriptome and epigenome are not just another measure of cell features — together, they represent the data sets that contain mechanistic though encrypted information of other cell phenotypes (including morphology, connectivity and physiology). To retrieve and interpret such information, we must obtain high-resolution (high-depth) omics data sets and, further, recognize and annotate the 'functional elements' and 'functional modules' embedded in these molecular profiles. This may be best illustrated by a comparison between the transcriptomic and epigenomic approach to cell typing versus the genomic approach to species classification.

In this postgenomic era, contemporary geneticists and evolutionary biologists routinely use genome alignment and comparative analysis to deduce the identity and evolutionary relationships of many organisms to enable classification and the generation of a taxonomy. However, this approach, now almost taken for granted, can succeed only if two requirements are met. First, high-resolution (that is, nucleotide) and high-coverage (preferably full genome) sequences are needed. Random subsets of genomes (for example, obtained with shot-gun sequencing at 10% coverage) are far less useful for genome comparison across species. Second, comparative analyses of genomic information must be guided by genetic principles and

mechanistic knowledge. Comparisons of, for example, GC content or overall genome-wide alignment are unlikely to be informative. The recognition of genomic functional elements such as the genetic code, exons, introns, promoters, enhancers and insulators, which constitute only a tiny fraction of the genomic landscape, is necessary to interpret and compare genome sequences, to extract biologically meaningful information, and to infer species identity and genetic (and evolutionary) relationships.

This genomic approach to species classification may apply to transcriptomic and epigenomic approaches to understanding the logic of cell-type diversity in individual organisms. First, high-resolution and high-coverage omics data sets are needed to provide the necessary substrates for comprehensive analysis and discovery. Large-scale, lower-resolution data sets alone (for example, assessing the expression on the order of 2,000 genes per cell[71–73]) may preclude the delineation of functional gene expression signatures. Second, we must delineate from these immense data sets 'functional transcriptomic signatures' (FTSs; for example, gene batteries[74,75] and cellular modules[28,49]) that build and customize molecular machines to shape cell phenotypes and properties. Furthermore, we must discover the biological principles that determine how genomic information encodes core transcriptional networks that orchestrate characteristic gene expression patterns in cardinal cell types. Therefore, just as the genomic approach has provided a unified framework and metric for species classification, transcriptomic and epigenomic analyses may provide a unified framework and metric for cell-type definition and classification.

Equipped with FTSs and an increasing understanding of the transcriptional algorithm of cardinal cell types, we may be able to make considerable strides in deciphering the logic of cell identity and diversity in multiple biomedical subdisciplines. For example, in the context of development, type-defining FTSs could provide the reference end points for exploring the developmental emergence and mechanism of cell identity[61,64]. In the context of evolution, where conserved cell types also acquire divergent features, core FTSs in mouse cardinal types could provide the essential axes with which true homologous cell types can be identified across species[76,77], more precisely than using sporadic and species-variable molecular markers. Indeed, such comparisons could further facilitate

the definition of core regulatory complexes (CoRCs) of transcription factors that have been suggested to enable the independent evolution of sister cell types and to regulate cell-type-specific traits[78]. In the context of reprogramming, where it has been difficult to distinguish whether in vitro differentiated cells have acquired a small number of measurable 'type-resembling features' (such as markers or measures of excitability and transmitter release) or a coherent cell-type identity[79–81], FTSs should provide a compelling and comprehensive evaluation of the reprogrammed cell status. In the context of disease, scRNAseq promises unprecedented cellular-resolution molecular diagnosis. However, when comparing single-cell transcriptomes from diseased samples with those from control samples, it is often difficult to determine whether differences in gene expression indicate a pathological state of a defined cell type in the disease sample or result from unintended comparison of different cell types between disease and control samples[82]. Comprehensive characterization of human type-defining FTSs would provide the necessary reference framework and precision to diagnose disease-related molecular alterations of specific cell types that affect tissue organization and function. At a fundamental level, such an explanatory framework may facilitate our understanding of the flow of biological information from the genomes, epigenomes and cell types, to tissue organization, function and behaviour, with implications in brain circuit evolution and disorders.

Z. Josh Huang [ID][1]* *and Anirban Paul* [ID][1,2]

[1]*Cold Spring Harbor Laboratory, New York, NY, USA.*

[2]*Department of Neural & Behavioral Sciences, Penn State University College of Medicine, Hershey, PA, USA.*

e-mail: huangj@cshl.edu

1. Bota, M. & Swanson, L. W. The neuron classification problem. *Brain Res. Rev.* **56**, 79–88 (2007).
2. Luo, L., Callaway, E. M. & Svoboda, K. Genetic dissection of neural circuits: a decade of progress. *Neuron* **98**, 865 (2018).
3. Roux, L. & Buzsaki, G. Tasks for inhibitory interneurons in intact brain circuits. *Neuropharmacology* **88**, 10–23 (2015).
4. Harris, K. D. & Shepherd, G. M. The neocortical circuit: themes and variations. *Nat. Neurosci.* **18**, 170–181 (2015).
5. Huang, Z. J. Toward a genetic dissection of cortical circuits in the mouse. *Neuron* **83**, 1284–1302 (2014).
6. Klausberger, T. & Somogyi, P. Neuronal diversity and temporal dynamics: the unity of hippocampal circuit operations. *Science* **321**, 53–57 (2008).
7. Tremblay, R., Lee, S. & Rudy, B. GABAergic interneurons in the neocortex: from cellular properties to circuits. *Neuron* **91**, 260–292 (2016).
8. Cajal, S. R. Nuevo concepto de la histología de los centros nerviosos. *Rev. Cienc. Med. Barc.* **18**, 361–376 (1892).
9. Kepecs, A. & Fishell, G. Interneuron cell types are fit to function. *Nature* **505**, 318–326 (2014).
10. Ascoli, G. A. et al. Petilla terminology: nomenclature of features of GABAergic interneurons of the cerebral cortex. *Nat. Rev. Neurosci.* **9**, 557–568 (2008).
11. DeFelipe, J. et al. New insights into the classification and nomenclature of cortical GABAergic interneurons. *Nat. Rev. Neurosci.* **14**, 202–216 (2013).
12. Hardcastle, K., Ganguli, S. & Giocomo, L. M. Cell types for our sense of location: where we are and where we are going. *Nat. Neurosci.* **20**, 1474–1482 (2017).
13. Harris, K. D. et al. Classes and continua of hippocampal CA1 inhibitory neurons revealed by single-cell transcriptomics. *PLOS Biol.* **16**, e2006387 (2018).
14. Zeng, H. & Sanes, J. R. Neuronal cell-type classification: challenges, opportunities and the path forward. *Nat. Rev. Neurosci.* **18**, 530–546 (2017).
15. Sugino, K. et al. Molecular taxonomy of major neuronal classes in the adult mouse forebrain. *Nat. Neurosci.* **9**, 99–107 (2006).
16. Zhang, Y. et al. An RNA-sequencing transcriptome and splicing database of glia, neurons, and vascular cells of the cerebral cortex. *J. Neurosci.* **34**, 11929–11947 (2014).
17. Eberwine, J. et al. Analysis of gene expression in single live neurons. *Proc. Natl Acad. Sci. USA* **89**, 3010–3014 (1992).
18. Tang, F. et al. mRNA-Seq whole-transcriptome analysis of a single cell. *Nat. Methods* **6**, 377–382 (2009).
19. Usoskin, D. et al. Unbiased classification of sensory neuron types by large-scale single-cell RNA sequencing. *Nat. Neurosci.* **18**, 145–153 (2015).
20. Macosko, E. Z. et al. Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* **161**, 1202–1214 (2015).
21. Zeisel, A. et al. Brain structure. Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. *Science* **347**, 1138–1142 (2015).
22. Tasic, B. et al. Adult mouse cortical cell taxonomy revealed by single cell transcriptomics. *Nat. Neurosci.* **19**, 335–346 (2016).
23. Tasic, B. et al. Shared and distinct transcriptomic cell types across neocortical areas. *Nature* **563**, 72–78 (2018).
24. Markram, H. et al. Reconstruction and simulation of neocortical microcircuitry. *Cell* **163**, 456–492 (2015).
25. Rudy, B. et al. Three groups of interneurons account for nearly 100% of neocortical GABAergic neurons. *Dev. Neurobiol.* **71**, 45–61 (2011).
26. Gerashchenko, D. et al. Identification of a population of sleep-active cerebral cortex neurons. *Proc. Natl Acad. Sci. USA* **105**, 10227–10232 (2008).
27. He, M. et al. Strategies and tools for combinatorial targeting of GABAergic neurons in mouse cerebral cortex. *Neuron* **91**, 1228–1243 (2016).
28. Paul, A. et al. Transcriptional architecture of synaptic communication delineates GABAergic neuron identity. *Cell* **171**, 522–539 (2017).
29. Obermayer, J. et al. Prefrontal cortical ChAT-VIP interneurons provide local excitation by cholinergic synaptic transmission and control attention. Preprint at *bioRxiv* https://www.biorxiv.org/content/10.1101/461723v1 (2018).
30. Tasic, B. Single cell transcriptomics in neuroscience: cell classification and beyond. *Curr. Opin. Neurobiol.* **50**, 242–249 (2018).
31. Nord, A. S. et al. Genomic perspectives of transcriptional regulation in forebrain development. *Neuron* **85**, 27–47 (2015).
32. Sussel, L. et al. Loss of Nkx2.1 homeobox gene function results in a ventral to dorsal molecular respecification within the basal telencephalon: evidence for a transformation of the pallidum into the striatum. *Development* **126**, 3359–3370 (1999).
33. Lim, L. et al. Development and functional diversification of cortical interneurons. *Neuron* **100**, 294–313 (2018).
34. Sandberg, M. et al. Transcriptional networks controlled by NKX2-1 in the development of forebrain GABAergic neurons. *Neuron* **91**, 1260–1275 (2016).
35. Vogt, D. et al. Lhx6 directly regulates Arx and CXCR7 to determine cortical interneuron fate and laminar position. *Neuron* **82**, 350–364 (2014).
36. Gouwens, N. W. et al. Classification of electrophysiological and morphological types in mouse visual cortex. *Nat. Neurosci.* (in the press).
37. Economo, M. N. et al. A platform for brain-wide imaging and reconstruction of individual neurons. *eLife* **5**, e10566 (2016).

38. Gong, H. et al. High-throughput dual-colour precision imaging for brain-wide connectome with cytoarchitectonic landmarks at the cellular level. *Nat. Commun.* **7**, 12142 (2016).

39. Wang, X. et al. Genetic single neuron anatomy reveals fine granularity of cortical axo-axonic cells. *Cell Rep.* **26**, 3145–3159 (2019).

40. Li, R. et al. Precise segmentation of densely interweaving neuron clusters using G-Cut. *Nat. Commun.* **10**, 1549 (2019).

41. Yang, J. et al. FMST: an automatic neuron tracing method based on fast marching and minimum spanning tree. *Neuroinformatics* **17**, 185–196 (2019).

42. Ascoli, G. A. & Wheeler, D. W. In search of a periodic table of the neurons: axonal-dendritic circuitry as the organizing principle: patterns of axons and dendrites within distinct anatomical parcels provide the blueprint for circuit-based neuronal classification. *Bioessays* **38**, 969–976 (2016).

43. Jiang, X. et al. Principles of connectivity among morphologically defined cell types in adult neocortex. *Science* **350**, aac9462 (2015).

44. Klausberger, T. et al. Brain-state- and cell-type-specific firing of hippocampal interneurons in vivo. *Nature* **421**, 844–848 (2003).

45. Varga, C., Golshani, P. & Soltesz, I. Frequency-invariant temporal ordering of interneuronal discharges during hippocampal oscillations in awake mice. *Proc. Natl Acad. Sci. USA* **109**, E2726–E2734 (2012).

46. Cadwell, C. R. et al. Electrophysiological, transcriptomic and morphologic profiling of single neurons using Patch-seq. *Nat. Biotechnol.* **34**, 199–203 (2016).

47. Fuzik, J. et al. Integration of electrophysiological recordings with single-cell RNA-seq data identifies neuronal subtypes. *Nat. Biotechnol.* **34**, 175–183 (2016).

48. Paul, A., Huang, Z. J. & Single-cell, R. N. A. sequencing of fluorescently labeled mouse neurons using manual sorting and double in vitro transcription with absolute counts sequencing (DIVA-Seq). *J. Vis. Exp.* **140**, e58690 (2018).

49. Hartwell, L. H. et al. From molecular to modular cell biology. *Nature* **402**, C47–C52 (1999).

50. Crow, M. et al. Characterizing the replicability of cell types defined by single cell RNA-sequencing data using MetaNeighbor. *Nat. Commun.* **9**, 884 (2018).

51. Lerner, T. N., Ye, L. & Deisseroth, K. Communication in neural circuits: tools, opportunities, and challenges. *Cell* **164**, 1136–1150 (2016).

52. DeFelipe, J. et al. A microcolumnar structure of monkey cerebral cortex revealed by immunocytochemical studies of double bouquet cell axons. *Neuroscience* **37**, 655–673 (1990).

53. Douglas, R. J. & Martin, K. A. Neuronal circuits of the neocortex. *Annu. Rev. Neurosci.* **27**, 419–451 (2004).

54. Szentagothai, J. The 'module-concept' in cerebral cortex architecture. *Brain Res.* **95**, 475–496 (1975).

55. da Costa, N. M. & Martin, K. A. Whose cortical column would that be? *Front. Neuroanat.* **4**, 16 (2010).

56. Sumbul, U. et al. Automated computation of arbor densities: a step toward identifying neuronal cell types. *Front. Neuroanat.* **8**, 139 (2014).

57. Helmstaedter, M. et al. Connectomic reconstruction of the inner plexiform layer in the mouse retina. *Nature* **500**, 168–174 (2013).

58. Jonas, E. & Kording, K. Automatic discovery of cell types and microcircuitry from neural connectomics. *eLife* **4**, e04250 (2015).

59. Kvitsiani, D. et al. Distinct behavioural and network correlates of two interneuron types in prefrontal cortex. *Nature* **498**, 363–366 (2013).

60. Pi, H. J. et al. Cortical interneurons that specialize in disinhibitory control. *Nature* **503**, 521–524 (2013).

61. Mayer, C. et al. Developmental diversification of cortical inhibitory interneurons. *Nature* **555**, 457–462 (2018).

62. Cao, J. et al. Joint profiling of chromatin accessibility and gene expression in thousands of single cells. *Science* **361**, 1380–1385 (2018).

63. Taniguchi, H., Lu, J. & Huang, Z. J. The spatial and temporal origin of chandelier cells in mouse neocortex. *Science* **339**, 70–74 (2013).

64. Mi, D. et al. Early emergence of cortical interneuron diversity in the mouse embryo. *Science* **360**, 81–85 (2018).

65. Hrvatin, S. et al. Single-cell analysis of experience-dependent transcriptomic states in the mouse visual cortex. *Nat. Neurosci.* **21**, 120–129 (2018).

66. Hu, H., Gan, J. & Jonas, P. Interneurons. fast-spiking, parvalbumin(+) GABAergic interneurons: from cellular design to microcircuit function. *Science* **345**, 1255263 (2014).

67. Seung, H. S. & Sumbul, U. Neuronal cell types and connectivity: lessons from the retina. *Neuron* **83**, 1262–1272 (2014).

68. Wagner, A., Regev, A. & Yosef, N. Revealing the vectors of cellular identity with single-cell genomics. *Nat. Biotechnol.* **34**, 1145–1160 (2016).

69. Shekhar, K. et al. Comprehensive classification of retinal bipolar neurons by single-cell transcriptomics. *Cell* **166**, 1308–1323 (2016).

70. Sabbagh, M. F. et al. Transcriptional and epigenomic landscapes of CNS and non-CNS vascular endothelial cells. *eLife* **7**, e36187 (2018).

71. Campbell, J. N. et al. A molecular census of arcuate hypothalamus and median eminence cell types. *Nat. Neurosci.* **20**, 484–496 (2017).

72. Saunders, A. et al. Molecular diversity and specializations among the cells of the adult mouse brain. *Cell* **174**, 1015–1030 (2018).

73. Zeisel, A. et al. Molecular architecture of the mouse nervous system. *Cell* **174**, 999–1014 (2018).

74. Hobert, O., Carrera, I. & Stefanakis, N. The molecular and gene regulatory signature of a neuron. *Trends Neurosci.* **33**, 435–445 (2010).

75. Thiebaud, P. et al. Intersecting batteries of differentially expressed genes in the early sea urchin embryo. *Genes Dev.* **4**, 1999–2010 (1990).

76. Hodge, R. D. et al. Conserved cell types with divergent features between human and mouse cortex. Preprint at *bioRxiv* https://www.biorxiv.org/content/10.1101/384826v1 (2018).

77. Tosches, M. A. et al. Evolution of pallium, hippocampus, and cortical cell types revealed by single-cell transcriptomics in reptiles. *Science* **360**, 881–888 (2018).

78. Arendt, D. et al. The origin and evolution of cell types. *Nat. Rev. Genet.* **17**, 744–757 (2016).

79. Birey, F. et al. Assembly of functionally integrated human forebrain spheroids. *Nature* **545**, 54–59 (2017).

80. Quadrato, G. et al. Cell diversity and network dynamics in photosensitive human brain organoids. *Nature* **545**, 48–53 (2017).

81. Tsunemoto, R. et al. Diverse reprogramming codes for neuronal identity. *Nature* **557**, 375–380 (2018).

82. Enwright Iii, J. F. et al. Transcriptome alterations of prefrontal cortical parvalbumin neurons in schizophrenia. *Mol. Psychiatry* **23**, 1606–1613 (2018).

83. Luo, C. et al. Single-cell methylomes identify neuronal subtypes and regulatory elements in mammalian cortex. *Science* **357**, 600–604 (2017).

84. Preissl, S. et al. Single-nucleus analysis of accessible chromatin in developing mouse forebrain reveals cell-type-specific transcriptional regulation. *Nat. Neurosci.* **21**, 432–439 (2018).

85. He, M. & Huang, Z. J. Genetic approaches to access cell types in mammalian nervous systems. *Curr. Opin. Neurobiol.* **50**, 109–118 (2018).

86. Schwarz, L. A. et al. Viral-genetic tracing of the input-output organization of a central noradrenaline circuit. *Nature* **524**, 88–92 (2015).

87. Kasthuri, N. et al. Saturated reconstruction of a volume of neocortex. *Cell* **162**, 648–661 (2015).

88. Coskun, A. F. & Cai, L. Dense transcript profiling in single cells by image correlation decoding. *Nat. Methods* **13**, 657–660 (2016).

89. Moffitt, J. R. et al. High-throughput single-cell gene-expression profiling with multiplexed error-robust fluorescence in situ hybridization. *Proc. Natl Acad. Sci. USA* **113**, 11046–11051 (2016).

90. Wang, X. et al. Three-dimensional intact-tissue sequencing of single-cell transcriptional states. *Science* **361**, eaat5691 (2018).

91. Eng, C. L. et al. Transcriptome-scale super-resolved imaging in tissues by RNA seqFISH. *Nature* **568**, 235–239 (2019).

92. Rodriques, S. G. et al. Slide-seq: a scalable technology for measuring genome-wide expression at high spatial resolution. *Science* **363**, 1463–1467 (2019).

Author contributions

Both authors researched data for the article and contributed substantially to the discussion of content. Z.J.H. wrote the article assisted by A.P. Both authors reviewed and edited the manuscript before submission.

Competing interests

The authors declare no competing interests.

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.