13 January 2019

Transcriptomic evidence for dense peptidergic neuromodulation networks in mouse cortex

Stephen J Smith, Uygar Sümbül, Lucas Graybuck, Forrest Collman, Sharmishtaa Seshamani, Rohan Gala, Olga Gliko, Leila Elabbady, Jeremy Miller, Trygve Bakken, Zizhen Yao, Ed Lein, Hongkui Zeng, Bosiljka Tasic, Michael Hawrylycz.

Allen Institute for Brain Science
615 Westlake Ave N, Seattle, WA 98109

Keywords: Neuropeptides, Mouse, Neocortex, Visual Cortex, Motor Cortex, Neuron Types, Neuron Taxonomy, mRNA-Seq, GPCR, Synaptic Plasticity, Neuromodulation, Homeostasis, Synaptic Network Homeostasis

Correspondence: Stephen J Smith, stephens@alleninstitute.org

Summary

Seeking new insights into intracortical neuromodulation, we have analyzed results from deep RNA-Seq analysis of 22,439 individual mouse neocortical neurons. With special interest in ways that cortical neurons may use paracrine neuropeptide (NP) signaling to modulate one anothers' synaptic and electrical function, we have concentrated on neocortical expression of genes encoding 18 neuropeptide precursor proteins (NPPs) and 29 cognate G-protein-coupled neuropeptide receptors (NP-GPCRs). We find that over 97% of sampled neurons express at least one of these NPP genes and that 98% express at least one NP-GPCR gene, with almost all neurons expressing several genes in each category. Reference to a transcriptomic taxonomy suggests that neocortical neuron types should be distinguishable by unique combinatorial signatures of NPP and NP-GPCR expression. We use these neuron-type-specific NP expression signatures to generate testable predictions regarding dense peptidergic neuromodulatory networks that may play prominent roles in cortical activity homeostasis and memory engram storage.

10,112 words (main text and figure legends), 120 references.

Introduction

Neuromodulation - the adjustment of synapse and ion channel function via diffusible cell-cell signaling molecules - is a fundamental requirement for all adaptive nervous system function across the entire animal kingdom [1–10]. Here we present evidence from single-cell transcriptomes for unexpectedly rich intracortical neuromodulation involving the evolutionarily ancient signaling molecules known as neuropeptides (NPs). Our findings indicate that all (or very nearly all) mouse cortical neurons express a variety of NP genes encoding both neuropeptide precursor proteins (NPPs) and neuropeptide-selective G-protein-coupled receptors (NP-GPCRs). NP gene expression is also found to be extraordinarily cell-type-specific and to involve a large palette of different NPP and NP-GPCR genes. The cell-type-specific patterning of NP gene expression predicts the existence of dense and highly multiplexed peptidergic signaling networks by which nearby cortical neurons may modulate one another's synaptic function and membrane excitability. The predicted neuropeptidergic networks are "dense" in the sense that they appear likely to involve almost every cortical neuron directly, both as a modulator source and a modulation target. Here we develop specific and testable hypotheses regarding basic roles played by neuropeptides in the homeostasis and plasticity of cortical synaptic circuitry.

Neuromodulatory signaling molecules take many different chemical forms, ranging from diatomic gases such as nitric oxide through amino acids, lipid and amino acid metabolites such as acetylcholine and dopamine to a variety of proteinaceous substances. By far the largest family of neuromodulatory molecules comprises the neuropeptides [11–16]. The most widely known and well-studied examples of neuropeptides are the endogenous "opioid" peptides - enkephalins, endorphins and dynorphins, encoded by four different NPP genes - but there are nearly one hundred other NPP genes in the human genome and numbers in all other known animal genomes are proportionately large.

The broadest definition of "neuropeptide" would embrace any soluble polypeptide that serves as a messenger by diffusing from one neuron to another. A narrower but more common definition [11] requires that (1) a neuropeptide precursor protein (NPP) be translated into the lumen of the source neuron's rough endoplasmic reticulum (rER) and (2) specifically cleaved during passage through the rER–Golgi complex, that (3) the cleavage product neuropeptide (NP) be packaged into dense-core vesicles (DCVs), transported and stored within the source neuron, then (4) released upon demand by activity- and calcium-dependent exocytosis, and only then (5) diffuse to act upon a target neuron by binding to a specific receptor. This pathway enlarges the palette of potential neuropeptide signals even beyond that established by the large number of NPP genes, as a given NPP can be cleaved into alternative NP products during DCV packaging and alternatively sorted during intracellular transport. Such alternative NP processing routes are cell-type-specific and result in NP products that activate NP receptors with differential biochemical and anatomical specificities.

Most neuropeptide receptors are members of the very large (many hundreds of genes) superfamily of G-protein-coupled receptor (GPCR) genes [12,17,18]. GPCRs are very selective high-affinity receptors distinguished by characteristic seven-transmembrane-segment structures and signal transduction involving heterotrimeric G-proteins (hence the name). Phylogenomic evidence suggests that early multicellular animals relied exclusively on peptides and GPCRs for the cell-cell signaling required for their slow and simple behaviors [17,19–22]. The later evolution of neurons, focal synaptic contacts, rapidly recycled small-molecule neurotransmitters, and ionotropic receptor-channels was presumably driven by survival advantages of the ever-faster signaling and behavior that these innovations enabled through the ages [19]. Though the fast synaptic transmission characteristic of contemporary higher animals is almost invariably based on recycling small molecule neurotransmitters and ionotropic receptors, modulation of synaptic transmission and membrane excitability by GPCRs remains very prominent across all contemporary animal phyla. Peptide roles appear to have evolved from cell-to-cell transmission to cell-type-to-cell-type modulation [17,19–22].

Though the receptors encoded by different GPCR genes have widely diverse peptide ligand specificities, there is considerable conservation at the level of cellular signal transduction mechanism. Most known GPCR actions on neurons reflect phosphorylation of ion channel or synaptic proteins, mediated by protein kinases dependent on the second messengers cyclic AMP and calcium. Effects of GPCR activation on

these second messengers, in turn, fall into just three major classes. The three classes are distinguishable by G-protein alpha subunit type class and corresponding second messenger impacts: The G-protein αi (I) class inhibits cAMP production, the Gαs class (S) stimulates cAMP production, and the Gαq class (Q) amplifies calcium signaling dynamics. For most GPCR genes, the primary G-protein α-subunit class (i.e., I, S or Q) is known and offers a worthy first-order prediction of the encoded GPCR's major signal transduction mode. This convergence to just three major GPCR transduction classes (I, S and Q) offers opportunities to simplify and generalize across GPCRs of widely diverse neuromodulatory ligand specificities. For instance, lessons from study of the actions of monoamine modulators like dopamine, serotonin and norepinephrine via monoamine-selective GPCRs on one synapse[23] are very likely to inform understanding of a neuropeptide-selective NP-GPCR of the same Gα class on another synapse. The profound functional consequences of GPCR modulation range from transformation of neuronal firing and calcium signaling dynamics through regulation of synaptic weights and short-term synaptic plasticity[1,6], to the gating of "Hebbian" or "Spike-Timing-Dependent" (STDP) rules governing long-term synaptic plasticity[24].

Because modulatory neuropeptides are not subject to the rapid transmitter re-uptake and/or degradation processes necessary for fast synaptic transmission, secreted neuropeptides persist long enough (e.g., minutes) in brain interstitial spaces for diffusion to NP-GPCRs hundreds of micrometers distant from release sites [16,25,26]. Neuropeptide signaling can thus be presumed "paracrine", with secretion from one cell acting upon many within a substantial diffusion distance and signals likewise converging from many local neurons onto one. The present analysis focuses on the mining of transcriptomic data for predictions of diffusive peptidergic signaling between neurons located near each other, within the volume confines of a single cortical local circuit (e.g., one "column", "barrel", or other such hypothesized small functional subunit of the cortical sheet).

It is well established that certain neuropeptides, including vasoactive intestinal peptide (VIP), somatostatin (SST), neuropeptide Y (NPY) and cholecystokinin (CCK), are present at high levels in specific subsets of cortical neurons [27]. These neuropeptides, consequently, have come into broad use as markers of particular classes GABAergic interneurons, while the corresponding NPP and NP-GPCR genetics have provided molecular access to those broad neuron-type classes [28,29]. *In situ* hybridization and microarray data (e.g., the Allen Brain Atlases [30,31]) have also established that mRNA transcripts encoding these four NPPs and that many other NPPs and cognate NP-GPCR genes are expressed in many brain regions. There has been a critical lack, however, of comprehensive expression data combining great genomic depth with single-cell resolution. Absent such data, it has been difficult to generate specific hypotheses regarding cortical neuropeptide function and to design experiments to test those hypotheses [12,27].

The present study is based on analysis of the single-cell RNA-Seq data on mouse cortical gene expression introduced in a 2018 publication by Tasic, et al. [32]. Tasic and co-workers acquired RNA-Seq data from a total of 22,439 individual neurons, with transcripts mapped to a depth of ~9,500 genes per cell (median) and a total of 21,931 protein-coding genes across all neurons sampled. Neurons were sampled from two distant and very different neocortical areas: 13,491 neurons from primary visual cortex (VISp), and 8,948 neurons from anterior lateral motor cortex (ALM). Tasic, et al., then demonstrate that their data supports robust clustering of neurons into more than one hundred transcriptomic neuron types. This data-driven taxonomy makes it possible to predict a protein "parts list" for a neuron of any given transcriptomic type. While additional work now under way [29,33–37] will be needed to reconcile these new transcriptomic taxonomies with existing anatomical and physiological neuron-type taxonomies, the transcriptomic taxonomy by itself already offers powerful new perspectives and testable predictions regarding the cell-type specificity of cortical neuromodulation.

In the following, we first describe new findings regarding cortical NPP and NP-GPCR gene expression that leverage simply the datasets' unprecedented single-cell resolution and genomic depth [32]. Then, we briefly introduce the 2018 Tasic transcriptomic neuron-type taxonomy and explore the further insights into neuropeptide signaling offered by a taxonomic framework [38]. Finally, we distill these findings into specific and testable predictions of local peptidergic modulation networks that may underlie the

homeostasis and plasticity of synapses and neuronal excitability necessary for all adaptive cortical circuit function.

Results

The present study focuses on expression at the mRNA level of genes encoding neuropeptide precursor and receptor proteins in individual neurons of mouse cortex. The acquisition and preliminary processing of transcriptomic data and development of a derivative neuron-type taxonomy are described in detail in the 2018 Tasic, et al. publication [32] and in additional narrative materials available on the Allen Brain Map website [39]. The complete 2018 Tasic transcriptomic datasets are available for download from the Allen Brain Atlas Data Portal website [40].

The 2018 Tasic data tables report counts-per-million-reads (CPM) measures of the abundance of transcripts mapping to a total of 21,931 protein-coding genes for all 22,439 individual neurons sampled. The CPM measure is a useful surrogate for transcript copy number, even though there may be some bias and imprecision when comparing transcripts of different lengths or when the number of discrete samples is small. Because the present analysis focuses primarily on very large differentials between cells and comparisons of same-gene transcripts across cells and cell types, and because very large numbers of cells are sampled, the limitations of the CPM measure should have little consequence for the analyses presented and conclusions drawn here.

The NP signaling genes upon which the present analysis focuses are all expressed very differentially across the sampled populations of individual mouse cortical neurons. That is, each gene is expressed at a high level in some subset of cells but at zero or very low levels in the remainder of the population. To compactly characterize such expression, we developed a "Peak CPM" metric. This metric is generated by ranking single-cell CPM values for a given gene across the entire population sampled, then designating the CPM value at the ascending 99.9th percentile as "Peak CPM". This metric was designed to minimize effects of sporadic outliers while yielding a good approximation to the actual peak expression of the specified gene within the highly expressing cell subset.

18 Neuropeptide Precursor Protein (NPP) genes are very highly expressed in mouse neocortex. **Table 1** lists 18 NPP genes highly expressed in varied subsets of individual neurons sampled from cortical areas VISp and ALM. This list was circumscribed by two requirements: (1) That the included NPP gene be highly expressed (~median or above Peak CPM for all protein-coding genes) in both VISp and ALM cortical areas, and (2) that at least one NP-GPCR gene cognate to a candidate NPP gene also be highly expressed in neurons within the same cortical local area. Requirement (2) was imposed here due to the focus here on paracrine peptidergic signaling as noted in **Introduction**. **Table 1** also lists Peak CPM values for each NPP gene, rank of that Peak CPM value as percentile across all protein-coding genes, predicted neuropeptide product(s) encoded, and the NP-GPCR gene(s) fulfilling requirement (2) for that NPP gene. Transcripts of no other known NPP genes met the criteria specified above.

The Peak CPM percentile column in **Table 1** shows that expression levels of most of the 18 NPP genes range are extremely high in the realm of Peak CPM values for all 21,931 protein-coding genes detected in all neurons sampled. Peak CPM values for three of these genes (Vip, Npy and Sst) are among the top 10 of all genes by single-cell peak CPM, five fall within the top 0.1%, and nine within the top 1%. The extremely high abundance of NPP transcripts suggests that NPP products are likely synthesized at correspondingly high rates. To maintain a steady state, the cell must therefore dispose of those products at a very high rate, with processing to active neuropeptides and secretion of those peptides being the most likely route of elimination.

	NPP Gene	Peak CPM	Peak CPM Percentile	Predicted Neuropeptides	Cognate NP-GPCR Genes
1	Vip	75,802	99.99%	Vasoactive Intestinal Peptide	Vipr1, Vipr2
2	Npy	59,909	99.98%	Neuropeptide Y	Npy1r, Npy2r, Npy5
3	Sst	42,191	99.97%	Somatostatins	Sstr1, Sstr2, Sstr3, Sstr4
4	Penk	19,519	99.92%	Enkephalins	Oprd1, Oprm1
5	Tac2	16,163	99.90%	Neurokinin B	Tacr3
6	Cck	14,002	99.87%	Cholecystokinins	Cckbr
7	Crh	12,036	99.82%	Corticotropin-Releasing Hormone	Crhr1, Crhr2
8	Tac1	9,228	99.70%	Substance P, Neurokinin A	Tacr1
9	Pdyn	7,164	99.58%	Dynorphins	Oprd1, Oprk1, Oprm1
10	Cort	3,724	98.92%	Cortistatin	Sstr1, Sstr2, Sstr3, Sstr4
11	Pthlh	2,504	98.08%	Parathyroid-Hormone-Like Hormone	Pth1r
12	Pnoc	1,581	96.13%	Nociceptins	Oprl1
13	Adcyap1	911	89.68%	Adenylate Cyclase-Activating Polypeptides	Adcyap1r1, Vipr1, Vipr2
14	Trh	706	83.03%	Thyrotropin-Releasing Hormone	Trhr, Trhr2
15	Grp	619	78.33%	Gastrin-Releasing Peptide	Grpr
16	Nmb	480	68.38%	Neuromedin B	Nmbr
17	Nts	295	52.65%	Neurotensin	Ntsr1, Ntsr2
18	Rln1	180	44.52%	Relaxin 1	Rxfp1, Rxfp2, Rxfp3

Table 1. 18 neuropeptide precursor protein (NPP) genes are highly expressed in mouse cortex. These genes are tabulated here along with peak single-cell expression levels as Peak CPM (see main text), percentile ranking of these Peak CPM values across Peak CPMs for all 21,931 protein-coding genes detected in all neurons sampled, predicted neuropeptide products, and genes encoding G-protein-coupled receptors (NP-GPCRs) that are cognate to the listed NPP genes and expressed cortex (see **Table 2**). Placement of NPP Peak CPMs on the value distribution for all protein-coding genes is represented graphically by **Supplementary Fig. 1A**. NPP genes are listed here in descending order of Peak CPM.

Expression of NPP genes by neocortical neurons is highly differential. **Figure 1A** characterizes the differential expression of the 18 NPP genes listed in **Table 1**. Each of 18 color-coded solid curves represents the distribution of single-neuron CPM values for one NPP gene. Curves were generated by plotting CPM for each individual neuron in descending order along a cell population percentile axis. Each curve shows an abrupt transition from very high to very low (commonly zero) expression across the sampled neuron population, but these transitions occur at very different population percentile points, providing clear evidence for highly differential single-cell expression of each gene. Percentages of the sampled neuron population expressing a given NPP gene (at greater than 1 CPM) range from more than 65% for Cck down to 1% for Nts.

All (or nearly all) neocortical neurons express at least one NPP gene. The dashed curve in the **Fig. 1A**, labeled "Max NPP Gene", was generated by plotting CPM values of the NPP gene with the highest CPM in each individual cell in descending order along a cell population percentile axis. This curve therefore shows that 97% percent of the sampled mouse cortical neurons express at least one NPP gene at >1 CPM and that 80% express at least one NPP gene at >1,000 CPM, a very high level. These numbers must be understood as lower limits to percentages of cortical neurons expressing at least one of the 18 NPP genes, taking into account the pulsatile nature of transcription [41] and the stochastic nature of RNA-Seq transcript sampling [42,43]. The results summarized in **Fig. 1A** may therefore be consistent with the proposition that every cortical neuron is peptidergic.

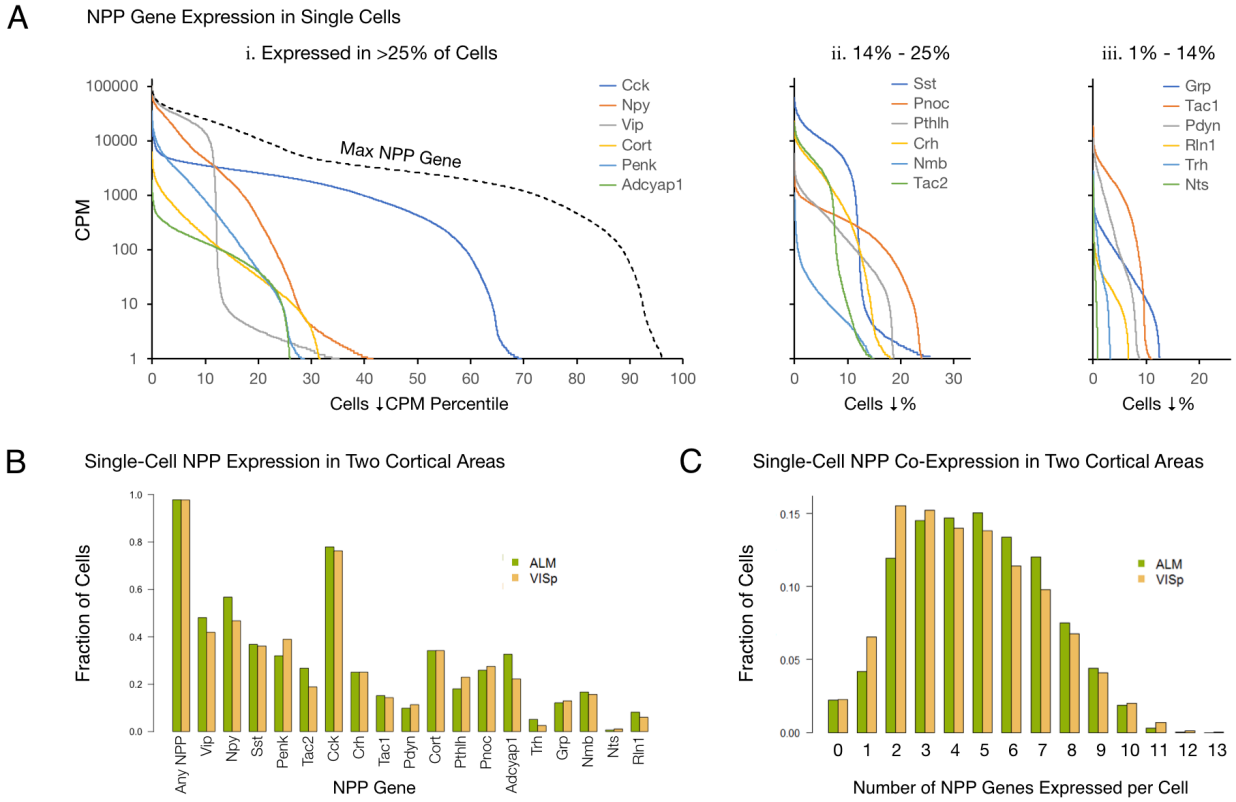


Figure 1. Expression and co-expression of NPP genes is highly differential with statistics conserved between two distant cortical areas. Single-cell expression analysis of 18 NPP genes in 22,439 individual neurons from cortical areas VISp and ALM. **(A)** Distributions of single-cell expression levels in neurons pooled from VISp and ALM. Color-coded solid curves plot single-cell CPM values for individual NPP genes in descending order along a cell population percentile axis. The 18 curves are segregated for clarity into three panels, sorted by cell population percentiles at which CPM values fall below 1. Large differences in fractions of cells expressing different NPP genes are evident. Dashed curve in left panel was generated by plotting CPM values of the highest CPM NPP gene for each individual cell in descending order. **(B)** Fractions of cells expressing each NPP genes represented separately for 13,491 VISp neurons and 8,948 ALM neurons, showing conservation between areas of the patterning of NPP expression fractions documented in panel A. **(C)** Histograms illustrating frequencies of various multiples of NPP gene co-expression in individual neurons, represented separately for VISp and ALM neurons. The paired vertical bars show strong conservation of co-expression patterns between the two areas.

Statistics of differential NPP gene expression are highly conserved between different neocortical areas. **Figure 1B** provides evidence for strong conservation of differential NPP expression profiles between VISp and ALM, two distant and very different neocortical areas. The paired bars represent fractions of cells expressing a given gene in each of the two areas. It is obvious that the differential expression profiles in VISp and ALM are highly similar, in spite of stark differences in function and cytoarchitecture between these two areas. Conservation of expression fractions across so many genes in such divergent cortical areas suggests that these patterns must have strong connections to conserved features of cortical function and argues against these patterns being secondary to more ephemeral variables such as neuronal activity patterns, which seem unlikely to be highly conserved between VISp and ALM areas.

Multiple NPP genes are co-expressed in almost all individual cortical neurons, with statistics conserved between divergent cortical areas. **Figure 1C** represents frequencies with which transcripts of various multiples drawn from the set of 18 NPP genes were detected in individual neurons. These data establish a remarkable degree of NPP gene co-expression in almost all individual cortical neurons. The modal number of co-expressed NPP genes detected is 2 in VISp and 5 in ALM, but both distributions are actually quite flat between 2 and 5, with broad plateaus out to 7 co-expressed NPP genes per cell and a substantial tail out to 10. The similarities of NPP co-expression distributions between the two distant neocortical areas suggests that NPP co-expression also has consequences that are conserved because they are fundamental to cortical function.

29 Neuropeptide-selective G-protein-coupled receptor (NP-GPCR) genes are highly expressed in mouse neocortex. **Table 2** lists 29 NP-GPCR genes that are highly expressed in subsets of individual neurons sampled from cortical areas VISp and ALM. These 29 genes encode receptor proteins selective for neuropeptide products encoded by the 18 NPP genes of **Table 1** (cross-referenced in that table as "Cognate NP-GPCR Genes"). **Table 2** provides quantitative information on expression levels of these 29 NP-GPCR genes, names the receptor proteins they encode, indicates the expected primary G-protein signal transduction type and cross-references the cognate cortically-expressed NPP genes. As noted above, the 18 NPP genes and 29 NP-GPCR genes listed in **Tables 1** and **2** were selected for focused analysis here due to their cognate pair relationships and the consequent prospect that they may constitute key signaling elements of intracortical peptidergic modulation networks.

	NP-GPCR Gene	Peak CPM	Peak CPM Percentile	Neuropeptide Receptor	Primary Gα Class	Cognate NPP Genes
1	Ntsr2	5,878	99.4%	Neurotensin Receptor 2	Q	Nts
2	Npy2r	1,248	94.3%	Neuropeptide Y Receptor Y2	I	Npy
3	Npy1r	1,087	92.8%	Neuropeptide Y Receptor Y1	I	Npy
4	Rxfp1	925	90.1%	Relaxin Family Receptor 1	S	Rln1
5	Adcyap1r1	902	89.5%	Adenylate Cyclase Activating Polypeptide 1 Receptor 1	S	Adcyap1, Vip
6	Sstr2	722	83.7%	Somatostatin Receptor 2	I	Sst, Cort
7	Cckbr	592	76.5%	Cholecystokinin B Receptor	Q	Cck
8	Grpr	577	75.4%	Gastrin Releasing Peptide Receptor	Q	Grp
9	Oprm1	556	74.0%	Mu-Opioid Receptor	I	Penk, Pdyn
10	Pth1r	499	69.8%	Parathyroid Hormone 1 Receptor	Q	Pthlh
11	Trhr	474	67.8%	Thyroid Releasing Hormone Receptor	Q	Trh
12	Nmbr	463	66.8%	Neuromedin B Receptor	Q	Nmb
13	Npy5r	439	64.8%	Neurpeptide Y Receptor Y5	I	Npy
14	Sstr1	403	61.8%	Somatostatin Receptor 1	I	Sst, Cort
15	Rxfp3	363	58.4%	Relaxin Family Receptor 3	S	Rln1
16	Tacr1	350	57.2%	Tachykinin Receptor 1	Q	Tac1
17	Oprl1	320	54.7%	Opioid Receptor-Like 1	I	Pnoc
18	Crhr1	319	54.7%	Corticotropin Releasing Hormone Receptor 1	S	Crh
19	Oprk1	302	53.2%	Kappa-Opioid Receptor	I	Pdyn
20	Tacr3	247	49.2%	Tachykinin Receptor 3	Q	Tac2
21	Crhr2	241	48.9%	Corticotropin Releasing Hormone Receptor 2	S	Crh
22	Vipr1	238	48.6%	Vasoactive Intestinal Peptide Receptor 1	S	Vip, Adcyap1
23	Vipr2	232	48.1%	Vasoactive Intestinal Peptide Receptor 2	S	Vip, Adcyap1
24	Rxfp2	177	44.3%	Relaxin Family Receptor 2	S	Rln1
25	Sstr4	148	42.2%	Somatostatin Receptor 4	I	Sst, Cort
26	Sstr3	135	41.5%	Somatostatin Receptor 3	I	Sst, Cort
27	Oprd1	111	39.6%	Delta-Opioid Receptor	I	Penk, Pdyn
28	Trhr2	103	38.9%	Thyroid Releasing Hormone Receptor 2	Q	Trh
29	Ntsr1	94	38.0%	Neurotensin Receptor 1	Q	Nts

Table 2. Genes encoding 29 NP-GPCRs expressed in mouse cortical areas VISp and ALM and cognate to the 18 neuropeptide precursor protein (NPP) genes listed in Table 1. Peak CPM values, Peak CPM percentile ranks and neuropeptide-selective GPCR protein encoded are tabulated for each NP-GPCR gene. In addition, the table lists the primary Gα signal transduction class associated with each predicted NP-GPCR as "I" for $G\alpha_{i/o}$ family, "S" for $G\alpha_s$ and "Q" for $G\alpha_{q/11}$ family and indicates cognate NPP genes. Placement of NP-GPCR Peak CPMs on the value distribution for all protein-coding genes is represented graphically by **Supplementary Fig. 1B**. NP-GPCR genes are listed here in order of Peak CPM.

While the high end of the range of cortical neuron Peak CPM values for NP-GPCR genes does not match the extreme values noted for NPP genes, Peak CPM values for NP-GPCR genes are nonetheless quite

high. The column labeled "Peak CPM Percentile" in **Table 2** shows that most NP-GPCR genes are expressed in cortex with Peak CPM values well above median (50th percentile) for all protein coding genes.

Expression of NP-GPCR genes by cortical neurons is highly differential. **Figure 2** represents expression patterns of the 29 NP-GPCR genes listed in **Table 2** in a manner that closely parallels the presentation for 18 NPP genes in **Fig.1**. **Figure 2A** establishes that each of the 29 NP-GPCR genes, like the 18 NPP genes, is expressed in highly differential fashion across the population of 22,439 mouse cortical neurons sampled. Each of 29 color-coded solid curves represents the distribution of single-neuron expression level values for one NP-GPCR gene. Curves were generated by plotting CPM for each individual neuron in descending order along a cell population percentile axis. As was noted for NPP genes in **Fig. 1**, each of the curves in **Fig. 2A** shows an abrupt transition from very high to very low (commonly zero) expression across the sampled neuron population. These transitions again occur at very different population percentile points, providing clear evidence for highly differential expression of NP-GPCR gene. Percentages of the sampled neuron population expressing a given NP-GPCR gene (at greater than 1 CPM) range from more than 72% for Adcyap1r1 down to 0.7% for Vipr2.

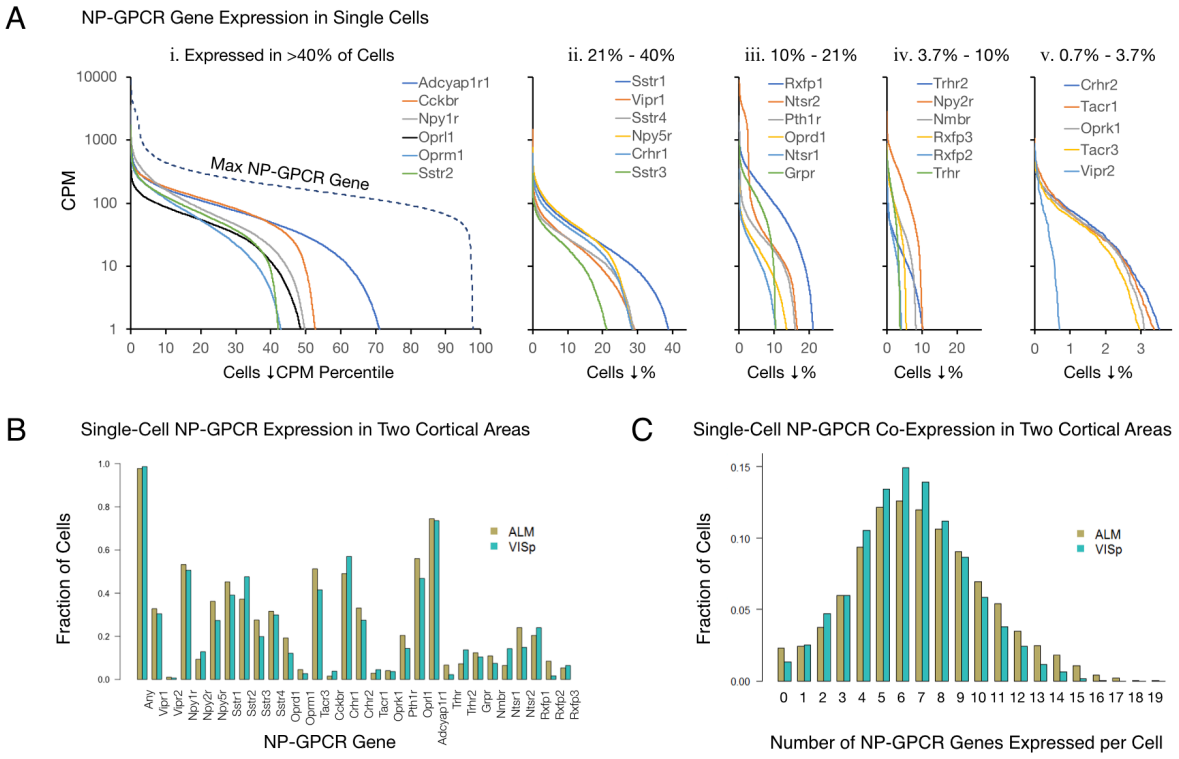


Figure 2. Expression and co-expression of NP-GPCR genes is highly differential with statistics conserved between two distant cortical areas. Single-cell expression analysis of 29 NP-GPCR genes in 22,439 individual neurons from cortical areas VISp and ALM. **(A)** Distributions of single-cell expression levels in neurons pooled from VISp and ALM. Color-coded solid curves plot single-cell CPM values for individual NP-GPCR genes in descending order along a cell population percentile axis. The 29 curves are segregated for clarity into five panels, sorted by cell population percentiles at which CPM values fall below 1. Large differences in fractions of cells expressing different NP-GPCR genes are evident. Dashed curve in left panel was generated by plotting CPM values of the highest CPM NP-GPCR gene for each individual cell in descending order. **(B)** Fractions of cells expressing each NP-GPCR genes represented separately for 13,491 VISp neurons and 8,948 ALM neurons, showing strong conservation between areas of the patterning of NP-GPCR expression fractions documented in panel A. **(C)** Histograms illustrating frequencies of various multiples of NP-GPCR gene co-expression in individual neurons, represented separately for VISp and ALM neurons. The paired vertical bars illustrate strong conservation of co-expression patterns between the two cortical areas.

Almost all (and possibly all) cortical neurons express at least one NP-GPCR gene. The dashed curve in the left panel of **Fig. 2A**, labeled "Max NP-GPCR Gene", was generated by plotting CPM values of the NP-GPCR gene with the highest CPM in each individual cell in descending order along a cell population percentile axis. This curve shows that 98% percent of the sampled mouse cortical neurons express at least one NP-GPCR gene at >1 CPM and that 78% express at least one NP-GPCR gene at >100 CPM,

lower than the comparable point for NPP genes (see **Fig. 1**) but still a very high value. Again, these numbers must be understood as lower limits to percentages of cortical neurons actually expressing at least one of the 29 NP-GPCR genes, after taking into account the pulsatile transcription and stochastic sampling issues cited above. The results summarized in **Fig. 2A** may thus be consistent with a conclusion that every cortical neuron expresses at least one NP-GPCR gene cognate to a cortically expressed NPP gene.

Statistics of differential NP-GPCR gene expression are highly conserved between different neocortical areas. **Figure 2B** provides evidence for strong conservation of differential NP-GPCR expression profiles between distant cortical areas VISp and ALM. The paired bars represent fractions of cells expressing a given gene in each of the two areas, revealing strong similarities of differential expression profiles in the two very different neocortical areas. Again, conservation of expression fractions across so many genes in such divergent cortical areas suggests that these patterns must have strong connections to conserved features of cortical function and thus argues against these patterns being secondary to more ephemeral variables such as neuronal activity patterns.

Multiple NP-GPCR genes are co-expressed in almost all individual cortical neurons, with statistics conserved between divergent cortical areas. **Figure 2C** represents frequencies of NP-GPCR gene co-expression multiples detected in individual neurons. These data establish that multiple NP-GPCR genes are co-expressed in almost all cortical neurons and that numbers of genes co-expressed are even higher than those noted above for co-expression of NPP genes. Modal numbers of co-expressed NP-GPCR genes detected is 6 in both VISp and ALM with broad plateaus extending out to 12 co-expressed NP-GPCR genes per cell. The striking similarities of NP-GPCR co-expression distributions between the two otherwise divergent neocortical areas once again suggests that the patterning of NP-GPCR co-expression may have consequences for cortical function that are conserved because they are functionally important.

Transcriptomic neuron-type taxonomy enables generation of testable predictions about neocortical neuropeptidergic signaling. Our analysis so far has relied solely upon the genomic depth and single-cell resolution characteristics of the 2018 Tasic transcriptomic data, without tapping into the powerful neuron-type taxonomy that was developed as a major goal of that study [32]. This 2018 Tasic taxonomy was developed from a large body of single-cell mRNA-Seq data based on dimensionality reduction and iterative hierarchical clustering methods [32]. The present analysis will make extensive use of a subset of the 2018 Tasic taxonomy representing 115 neurons types discriminated in VISp and ALM cortical areas, as summarized in **Supplementary Fig. 2**. This taxonomy will be represented in the following figures by cladograms and/or cell-type color codes that can be interpreted by reference to **Supplementary Fig. 2** or the 2018 Tasic publication [32].

Expression of the 18 NPP genes is very highly neuron-type specific. **Figure 3A** represents expression levels of the 18 NPP genes across all 115 VISp+ALM neuron types as a "heat map" matrix color coding $\log_{10}$ CPM values for each NPP and each neuron type. The CPM values so rendered are calculated as "trimmed means" (mean value after discarding the top 1% of distributions to suppress outlier effects) of single-cell CPM values aggregated by each neuron-type cluster (commonly on the order of 100 cells, see **Supplementary Figs. 2A** and **2C** for actual cell counts). **Figure 3A** confirms and extends four reasonable expectations from the type-agnostic single-cell analyses of **Figs. 1 and 2** above: (1) neurons of every type express one or more of the 18 NPP genes, (2) each of the 18 NPP genes is expressed in multiple neuron types, (3) neurons of every type express multiple NPP genes, and (4) expression of NPP genes is highly differential across neuron types. Remarkably, **Fig. 3A** shows that type-to-type variations in expression level for every one of the 18 NPP genes span the full >10,000-fold dynamic range characteristic of the Tasic 2018 RNA-Seq data. Quite intriguingly, **Fig. 3A** also suggests that each of the 115 VISp+ALM cell types might be distinguished by a unique combinatorial pattern of NPP gene expression. This possibility will be explored quantitatively in connection with **Fig. 4** below.

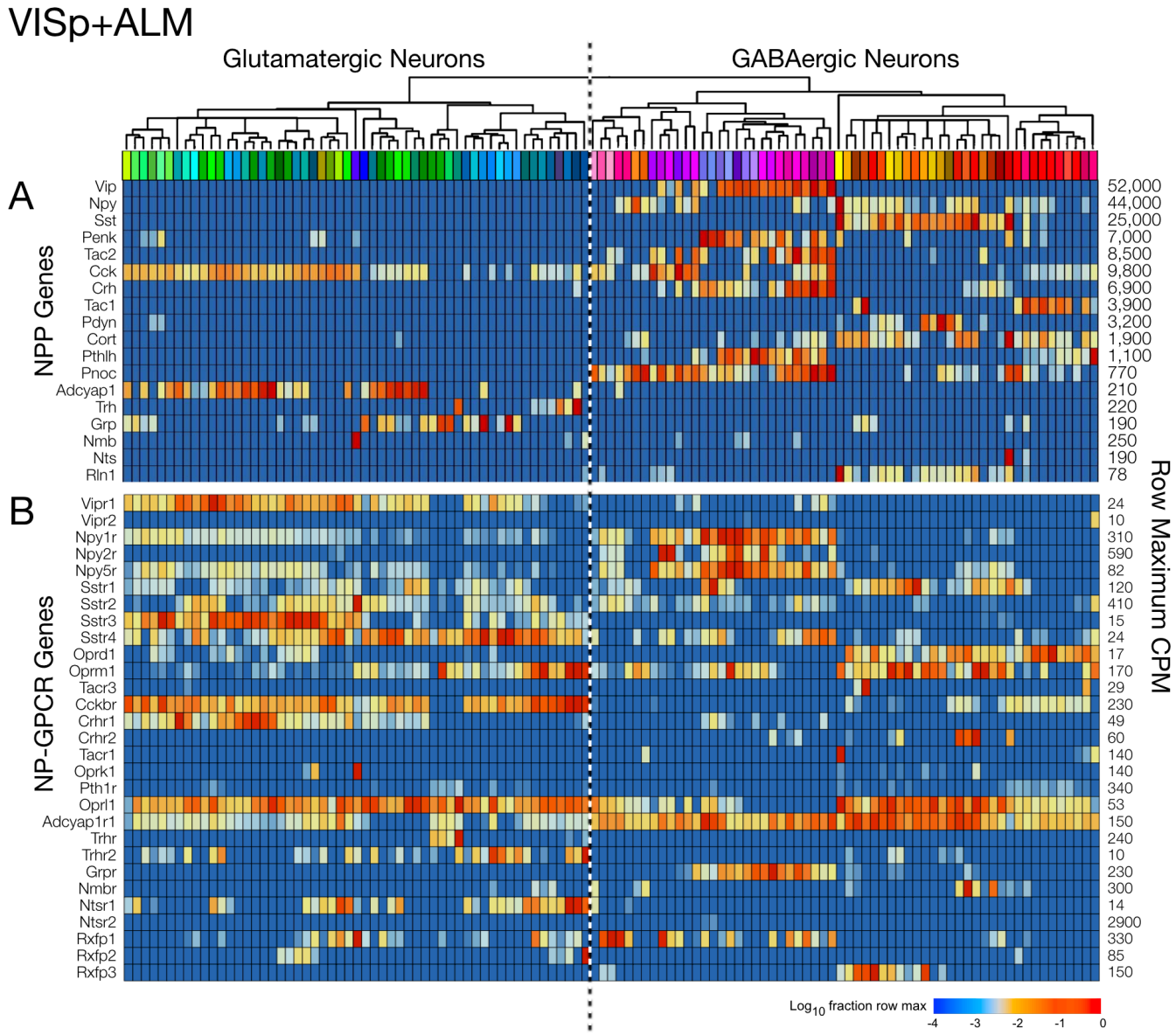


Figure 3. Profiles across 115 transcriptomic neuron types reveal very high type specificity of NP gene expression. Heat maps representing CPM expression values of 18 NPP and 29 NP-GPCR genes (rows) aggregated for each of 115 neuron types (columns) distinguished by the 2018 Tasic taxonomy [32]. Color mapping (key at lower left) encodes $\log_{10}$ mean of CPM values for each gene and cell type after trimming away top 1% percent CPM values to suppress outlier effects. Expression values are normalized separately for each gene (row), with numbers at right representing row maximum trimmed mean CPM values. Neuron-type taxonomic cladogram and color scheme at top of this map are those introduced in the 2018 Tasic, et al., publication and summarized in the present **Supplementary Fig. 2**. Note that every gene spans a very wide range of expression values across all neuron types. **(A)** Expression profiles of 18 NPP genes, ordered as in Table 1. **(B)** Expression profiles of 29 NP-GPCR genes for all VISp and ALM neuron types, with NP-GPCR genes ordered as in **Table 2.**

Figure 3A provides for ready comparison of NPP gene expression patterns between glutamatergic and GABAergic cell types. Clearly, GABAergic types are clearly more prolific in the variety and strength of their NPP genes expression. While glutamatergic types express fewer NPP genes, and do not match the extremely high NPP expression levels observed in almost every GABAergic type, each type nonetheless expresses at least one, and generally more, NPP genes at a very substantial level. This differential is consistent with a long history of neuroscientific use of neuropeptide products as protein markers of GABAergic neuron subsets (e.g., VIP, SST, NPY), which has no parallel in distinguishing glutamatergic neuron subsets.

Expression of the 29 NP-GPCR genes is highly neuron-type specific. **Figure 3B** illustrates neuron type specificity of NP-GPCR expression in a manner identical to the treatment of NPP gene expression in **Fig.**

3A and invites analogous conclusions: (1) neurons of every neuron type express one or more of the 29 NP-GPCR genes at very high levels, (2) neurons of every type express multiple NP-GPCR genes, and (3) expression of NP-GPCR genes is highly differential across neuron types. **Figure 3B** also shows, however, that the stronger and more varied expression of NPP genes in GABAergic expression profiles that was evident in **Fig. 3A** is leveled or even reversed for NP-GPCR genes. That is, while GABAergic neurons clearly show the more prolific and varied expression of NPP genes, glutamatergic neurons may be somewhat more prolific expressors of NP-GPCR genes.

A transcriptomic signature based upon just 47 NP-signaling genes (18 NPP, 29 NP-GPCR) suffices for accurate classification of neocortical neurons. The strong marker patterning of the 47 NP gene expression profiles evident in **Fig. 3** suggests the possibility that each of the 115 neuron types profiled in that figure might be distinguished by a unique combination of these 18 NPP and 29 NP-GPCR genes. To explore this possibility quantitatively, we formulated the problem by asking whether there exists a low dimensional representation of gene expression that naturally separates neurons of different types into distinct parts of that low-dimensional space. The extent to which a cell's location in such a space can be inferred from the expression of a limited subset of genes (such as our 47 NP genes) should then provide a measure of the sufficiency of that subset to classify a cell accurately. Hierarchical clustering methods to define cell types based upon gene expression are well established [44,45] but have difficulty when comparing and making inferences between datasets. We therefore devised a machine learning approach based on linked multi-layer autoencoders to address this question explicitly and quantitatively.

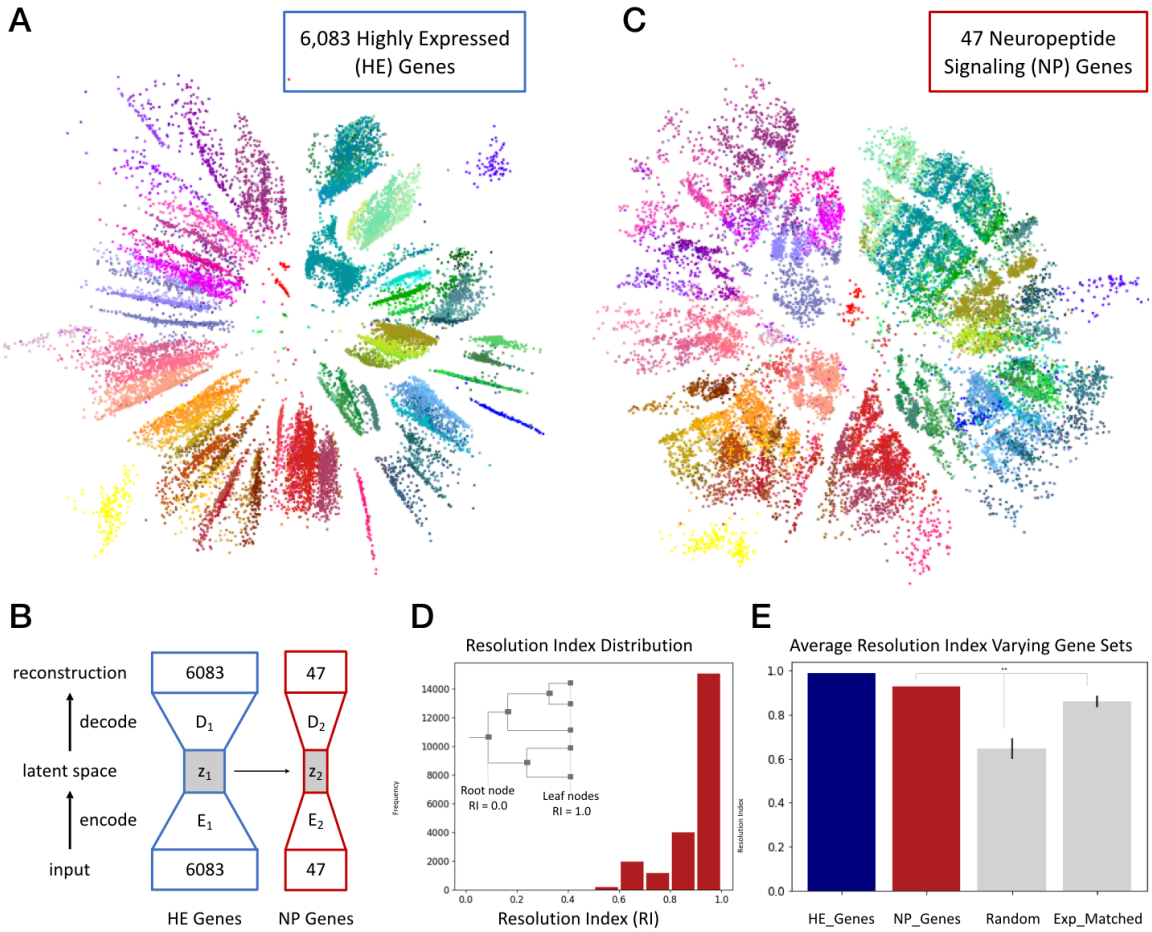


Figure 4. Neurons can be classified effectively based on just 47 NP genes (18 NPP and 29 NP-GPCR) alone. (A) A two-dimensional latent space representation of autoencoder network neuron-type classification of 22,439 cells based on 6,083 highly expressed (HE) genes, where each dot represents a single cell colored according to the 2018 Tasic [32] type-code coloring scheme, as introduced here in **Fig. 3** and **Supplementary Fig. 2**. The clustering of type-code colors to form distinct islands is indicative of good clustering performance. **(B)** Schematic representation of the linked autoencoder architecture used to optimize a second network to classify neurons as similarly as possible into a latent space based on much smaller gene

sets. **(C)** Representation of latent space z_2 **(D)** A Gaussian mixture model of cell types is fit to a five-dimensional NP gene representations using 2018 Tasic classification. Progressively simplified taxonomies are obtained by iterative merging, and the most complex taxonomy for which each cell is correctly mapped is recorded as that cell's Resolution Index (RI). This histogram represents the distribution of RI scores corresponding to classification of the five-dimensional NP gene latent space representation. **(E)** Average RI scores for latent space representations encoded by various gene subsets. The average RI per cell obtained from the NP genes representation (0.928) was significantly higher than the average obtained from 100 distinct sets of 47 random genes (0.645), and the average of 100 sets of 47 genes that matched expression levels of the 47 NP genes (0.858), $p<0.01$. See **Supplementary Methods** for additional details of autoencoder development and performance metrics.

A single autoencoder network was first developed and trained to encode the expression of a very large set of 6,083 highly expressed neuronal genes (the "HE" gene set) the gene expression based upon the 22,439 neurons in Tasic 2018. Results are illustrated in **Figure 4A**, where encoding coordinates in a two-dimensional latent space of 22,439 individual cells are displayed as discrete dots colored according to their prior Tasic 2018 hierarchical classification (the neuron-type color code introduced here in **Fig. 3** and **Supplementary Fig. 2**). The tight grouping of type-code colors evident in **Fig. 4A** implicitly represents that position within this latest space corresponds well to the neuron types defined by hierarchical classification, in spite of the fact that the autoencoder was given no explicit information about how neurons were classified in Tasic 2018. We then trained a second, linked auto-encoder with the architecture schematized in **Fig. 4B** to classify cells using only the 47-gene subset, under a cost function constraint that latent spaces of the two auto-encoders be as similar as possible. This allowed us to test the extent to which any small gene subset by itself could match the encoding performance obtained using the much larger gene set. **Fig. 4C** displays a two-dimensional latent space resulting from encoding the same 22,439 neurons based only on 47 NP genes tabulated above, again projecting one dot for each cell using the Tasic 2018 type-code colors. Once again, the tight color grouping evident in **Fig. 4C** suggests qualitatively that these 47 genes indeed enable excellent type encoding of individual neurons.

For quantitative comparison of classification performance based on varied neocortical gene sets, we partitioned the autoencoder encodings into classes using a supervised Gaussian Mixture Model (see methods) and designed the resolution index schematized in **Fig. 4D** to evaluate consensus between classifications driven by autoencoder encoding with the 2018 Tasic hierarchical taxonomy of reference. This index yields a value of 0 when a cell is mapped incorrectly from the root node and 1 when a cell is mapped correctly all the way to a terminal leaf node. By averaging this metric over all 22,439 neurons, we generated an overall figure of merit called a resolution index. This figure for the large HE 6083 gene set was 0.987, the same index for classification based on the NP 47-gene subset was 0.928. To place these resolution index numbers in context and test the significance of this correspondence, we compared resolution indices resulting from linked autoencoder classification based on 100 subsets of 47 genes drawn randomly from the Tasic 2018 expression dataset. The sets of 47 random genes yielded an average resolution index of 0.645 ± 0.047 (Fig. 4E), establishing clearly that NP genes yield classification greatly superior to random subsets of 47 genes. **Figure 4E** also shows results from encoding runs using 100 sets of 47 random genes chosen to approximate the same high expression statistics of the NP genes. Again, resolution indices from the random sets fell well below that yielded by the 47 NP genes (average $= 0.858 \pm 0.0242$, with none reaching the NP gene index of 0.928 and the difference being significant at $p>0.01$). These results thus reinforce earlier indications of some especially close and fundamental connection between neuropeptide gene expression and neuronal cell-type identity.

Cell-type-specificity of differential NP gene expression is conserved between neocortical areas. **Figure 5** juxtaposes separate VISp and ALM expression profiles for NPP and NP-GPCR genes across 93 VISp neuron types (**Fig. 5A**) and 84 ALM neuron types (**Fig. 5B**). Similarities of expression profiles for the two areas are obvious in **Fig. 5**, but there are also visible differences. The latter are rooted primarily in the substantial divergence of glutamatergic neuron taxonomies discussed at length in Tasic, et al. [32] and summarized here in **Supplementary Fig. 3**. Very strong similarities of both NPP and NP-GPCR expression profiles are most obvious for the GABAergic types, where the taxonomies are identical except for the absence of two GABAergic types in ALM (indicated by gray vertical placeholder bars in **Fig. 5B**). The general conservation of neuron-type expression patterns between the two distant neocortical areas (NPP correlation: $\rho= 0.974$, $p<2.2e-16$, NP-GPCR: 0.877, $p<2.2e-16$) thus provides yet another indication of robust connection between NP gene expression and cortical neuron differentiation.

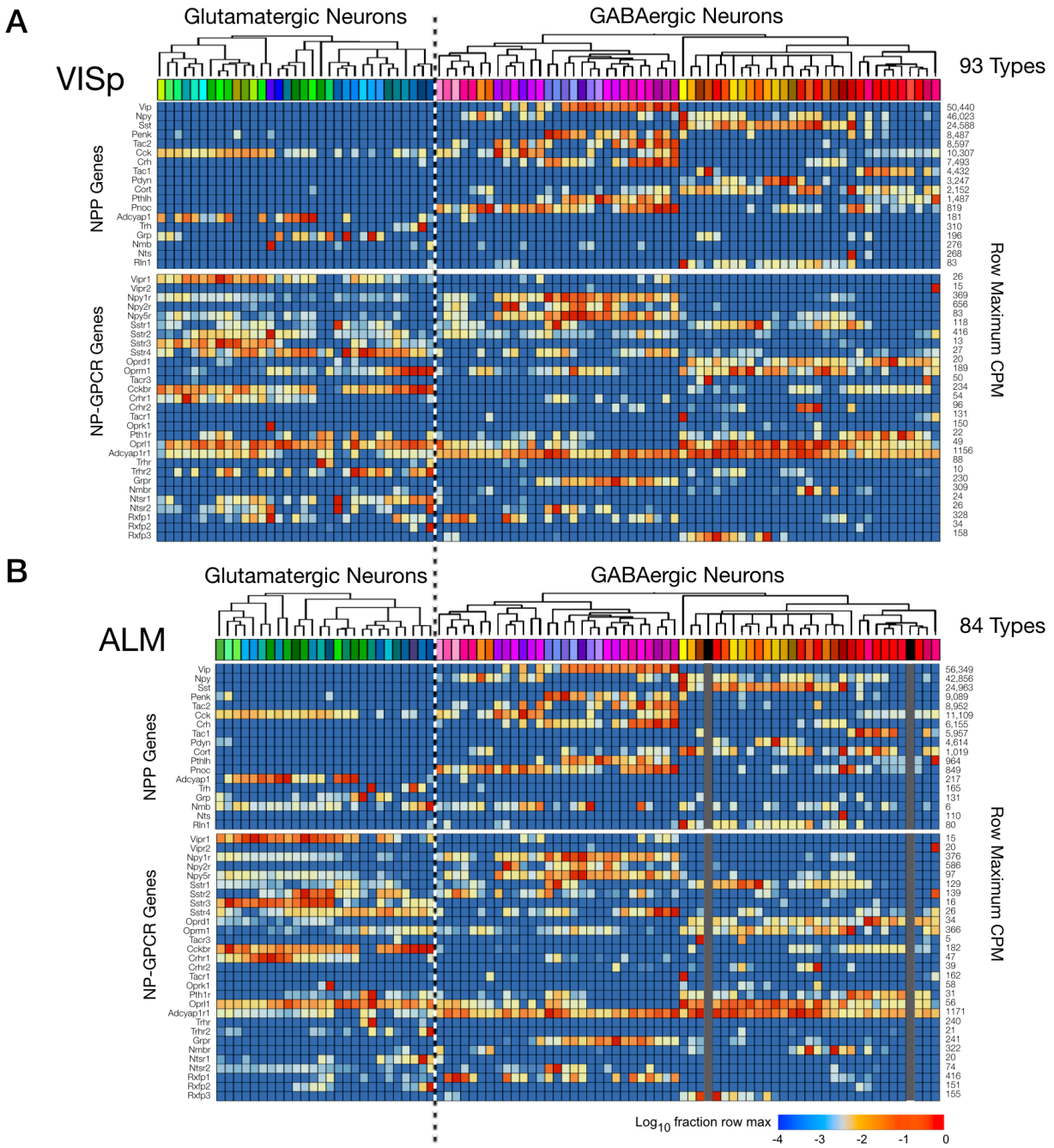


Figure 5. Expression profiles of NP genes restricted to cortical area VISp or to area ALM are very similar but differ somewhat due to divergence of VISp and ALM neuron-type taxonomies. Neuron-type-based NPP and NP-GPCR expression heat maps similar to those of Fig. 3 but separating data from area VISp and ALM samples. Heat maps are placed here to vertically align GABAergic neuron types that match between VISp and ALM areas. Vertical gray bars in Fig. 5B are spacers marking the two GABAergic cell types absent in ALM. Glutamatergic taxonomies differ more substantially (see Tasic, et al. [32] and Supplementary Fig. 3). **(A)** Expression profiles for 18 NPP and 29 NP-GPCR genes in area VISp, formulated, color coded and ordered as in Fig. 3. **(B)** Expression profiles for 18 NPP and 29 NP-GPCR genes in area ALM, again displayed as in Fig. 3.

Expression of 37 cognate NPP/NP-GPCR pairs in cortex predicts the potential existence of 37 intracortical peptidergic networks. Expression of an NPP gene in one neuron and a cognate NP-GPCR gene in another nearby neuron implies a prospect of local signaling, with secretion of a specific neuropeptide by the first neuron activating a specific cognate modulatory receptor on the second neuron. The present set of 47 cortical NP genes (18 NPP and 29 NP-GPCR) comprises the 37 distinct cognate NPP/NP-GPCR pairs enumerated in **Table 3** and accordingly predicts at least 37 distinct peptidergic neuromodulation networks. ("At least" because the precursor protein product of a single NPP gene may

be cleaved alternatively in different neuron types to produce alternative active neuropeptides in different neuron types and thus grow the number of potentially distinct peptidergic networks.) As noted in the **Introduction**, expected neuropeptide diffusion distances suggest that any neuron within a local cortical area (e.g., VISp or ALM) might signal by diffusion to any other neuron within that same local area, but almost surely not to more distant areas (e.g., from VISp to ALM). In the following, we'll therefore make predictions of 37 x 2 peptidergic signaling networks, keeping separate consideration of signaling within VISp and within ALM.

#	Cognate Pair Symbol	NPP Gene	NP-GPCR Gene	Primary Gα Class	Fraction of Pairs
1	Vip→Vipr1	Vip	Vipr1	S	0.140
2	Vip→Vipr2	Vip	Vipr2	S	0.004
3	Npy→Npy1r	Npy	Npy1r	I	0.263
4	Npy→Npy2r	Npy	Npy2r	I	0.057
5	Npy→Npy5r	Npy	Npy5r	I	0.159
6	Sst→Sstr1	Sst	Sstr1	I	0.151
7	Sst→Sstr2	Sst	Sstr2	I	0.158
8	Sst→Sstr3	Sst	Sstr3	I	0.084
9	Sst→Sstr4	Sst	Sstr4	I	0.111
10	Penk→Oprd1	Penk	Oprd1	I	0.053
11	Penk→Oprm1	Penk	Oprm1	I	0.163
12	Tac2→Tacr3	Tac2	Tacr3	Q	0.006
13	Cck→Cckbr	Cck	Cckbr	Q	0.414
14	Crh→Crhr1	Crh	Crhr1	S	0.075
15	Crh→Crhr2	Crh	Crhr2	S	0.010
16	Tac1→Tacr1	Tac1	Tacr1	Q	0.006
17	Pdyn→Oprd1	Pdyn	Oprd1	I	0.016
18	Pdyn→Oprk1	Pdyn	Oprk1	I	0.004
19	Pdyn→Oprm1	Pdyn	Oprm1	I	0.049
20	Cort→Sstr1	Cort	Sstr1	I	0.142
21	Cort→Sstr2	Cort	Sstr2	I	0.148
22	Cort→Sstr3	Cort	Sstr3	I	0.079
23	Cort→Sstr4	Cort	Sstr4	I	0.105
24	Pthlh→Pth1r	Pthlh	Pth1r	Q	0.035
25	Pnoc→Oprl1	Pnoc	Oprl1	I	0.135
26	Adcyap1→Adcyap1r	Adcyap1	Adcyap1r1	S	0.196
27	Adcyap1→Vipr1	Adcyap1	Vipr1	S	0.084
28	Adcyap1→Vipr2	Adcyap1	Vipr2	S	0.002
29	Trh→Trhr	Trh	Trhr	Q	0.002
30	Trh→Trhr2	Trh	Trhr2	Q	0.004
31	Grp→Grpr	Grp	Grpr	Q	0.014
32	Nmb→Nmbr	Nmb	Nmbr	Q	0.014
33	Nts→Ntsr1	Nts	Ntsr1	Q	0.001
34	Nts→Ntsr2	Nts	Ntsr2	Q	0.002
35	Rln1→Rxfp1	Rln1	Rxfp1	S	0.015
36	Rln1→Rxfp2	Rln1	Rxfp2	S	0.003
37	Rln1→Rxfp3	Rln1	Rxfp3	I	0.004

Table 3. The 18 NPP and 29 NP-GPCR genes of Tables 1 and 2 comprise 37 cognate NPP/NP-GPCR pairs and predict at least 37 potentially distinct peptidergic modulatory networks. The 37 pairs are enumerated here along with indications of the expected primary GPCR signal transduction class for each NP-GPCR [46] and an indication of the gene expression prevalence for each cognate pair as a fraction of the total number of neuron pairs surveyed.

Type-specific NP gene expression profiles predict type-specific peptidergic coupling. Because NP-GPCRs are highly selective for the neuropeptide products of specific NPP genes, as indicated in **Table 3**, the predicted neuropeptide networks may transmit 37 or more distinct and independent modulatory signals**. Figure 6** displays heat maps representing predictions of peptidergic coupling from a selection of the 37 cognate NP gene pairs and expression profiles of the paired NPP and NP-GPCR genes. The predictions of **Fig. 6** are based on cell-type-by-cell-type aggregation of binarized cell-pair-by-cell-pair products of the NPP and NP-GPCR gene CPM values. CPM values were first thresholded at the 50th percentile (t_p, t_q) independently for each cell type. The coupling matrix $E(p,q)$ is then defined as:

$$E(p,q) = \frac{1}{|C(p)||C(q)|} \sum_{j=1}^{|C(p)||C(q)|} I(e^j{}_p > t_p) * I(e^j{}_q > t_q) \qquad (1)$$

where *ej*p denotes the expression of individual cell j in cell type p, |C(p)| denotes the total number of expressing cells of type p, and I is the indicator function. $E(p,q)$ is therefore the fraction of expressing pairs exceeding the 50th percentile threshold.

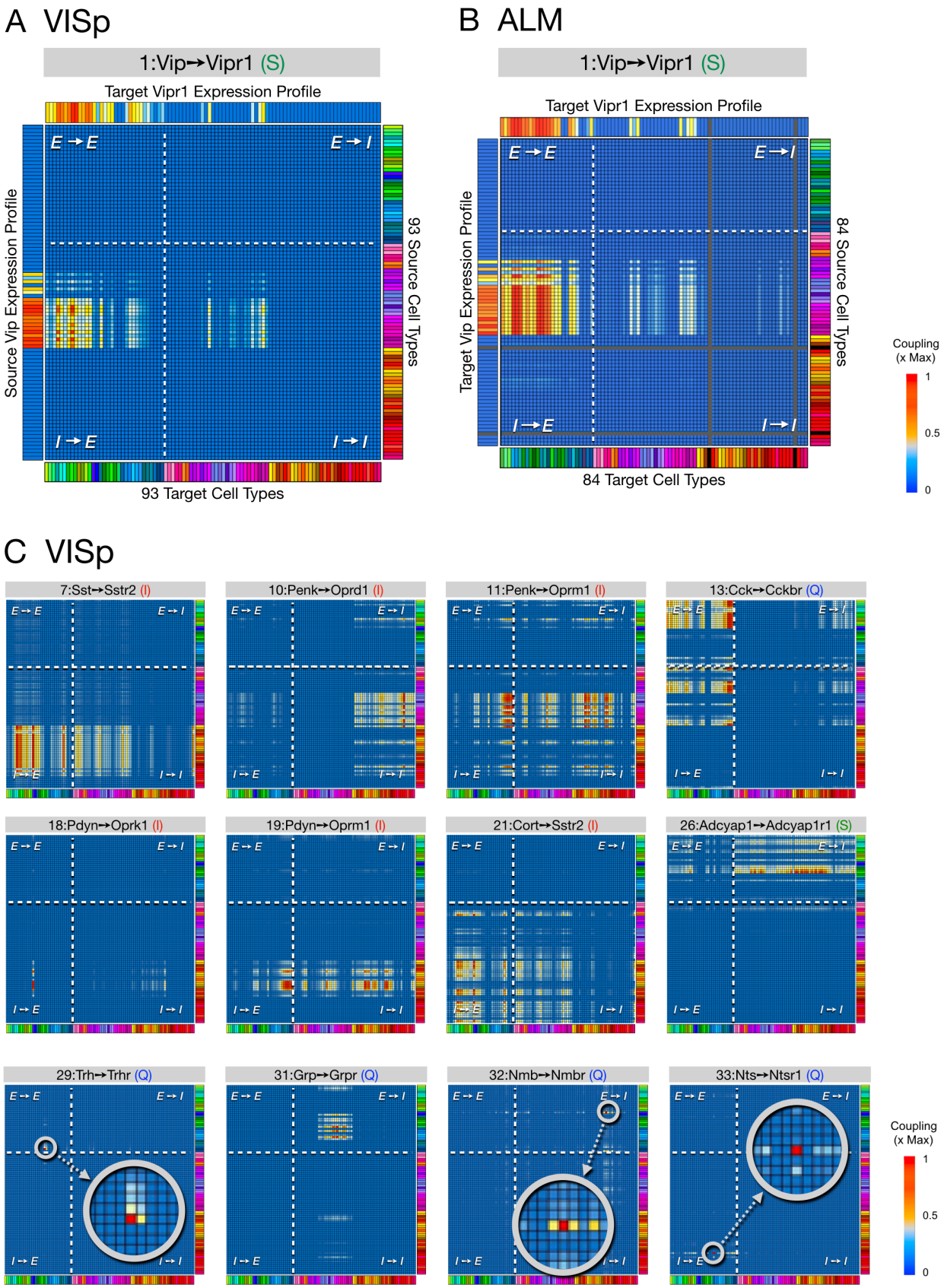


Figure 6. Expression profiles of conjugate NPP/NP-GPCR pairs predict neuropeptidergic signaling networks. Matrix heat maps represent predictions of local peptidergic coupling between specific neuron types based on analysis of the 2018 Tasic dataset using Eqn.1. Coupling strength (relative to matrix maximum) is represented as indicated by color scale keys and for each cell-type pair as indicated by type-code color strips at bottom and right margins of each map. Dashed cross on each

plot demarcates quadrants of glutamatergic-to-glutamatergic neuron coupling (*E→E*), glutamatergic- glutamatergic-to-GABAergic coupling (*E→I*), GABAergic-to-GABAergic coupling (*I→I*) and GABAergic-to-glutamatergic coupling (*I→E*). **(A)** An exemplar coupling matrix based on area VISp neuron-type expression profiles for the cognate pair of Vip and Vipr1, which are displayed as profile strips on the matrix plot's left and top margins. The Ga signal transduction class for the NP-GPCR (Vipr1 in this case) is designated here and in all subsequent panels by the parenthetical alphabetic character. **(B)** Exemplar matrix for Vip-Vipr1 coupling in area ALM, similar to VISp treatment in (A). Vertical black bars in type-code color strip and gray bars in heat map here are alignment-preserving spacers marking the two GABAergic cell types absent in ALM. Glutamatergic type alignments differ more substantially (see Tasic, et al. [32] and Supplementary Fig. 3). **(C)** Twelve of the 37 coupling matrices predicted for area VISp, selected to exemplify the wide range of coupling motifs predicted from the wide range of type-specific expression profiles evident in Fig. 5A. The "magnifying glass" insets in panels 29, 32 and 33 highlight narrow coupling, with putative network 29 (Trh→Trhr) predicting significant coupling only from neuron type *L6b VISp Crh* to *L5 NP VISp Trhr Cpne,* 32 (Nmb→Nmbr) from *L6 IT VISp Car3* to *Sst Crh 4930553C11Rik*, and 33 (Nts→Ntsr1) from *Sst Nts* to *L6 CT Nxph2 Sla*

The exemplar matrix displayed in **Fig. 6A** predicts coupling in area VISp based on VISp expression profiles of the *1:Vip→Vipr1* cognate pair. **Figure 6B** represents a similar prediction for the same pair in area ALM. The dashed white crosses overlying both plots partition the matrices based on pairings of glutamatergic and GABAergic neuron types. Major features of the VISp and ALM coupling matrices are highly similar. Both matrices predict strong signaling from the canonical broad class of VIP-positive GABAergic neurons to a broad subset of glutamatergic neurons that strongly express the Vipr1 NP-GPCR: the strongest coupling thus falls in the *I→E* quadrant. Weaker coupling is observed in the *I→I* quadrant, where the Vipr1 NP-GPCR gene is less strongly expressed in a few GABAergic cell types. Apparent differences between VISp and ALM coupling predictions are mainly due to exclusive expression of different glutamatergic cell types in the two areas, and only in small part due to difference in same-type expression within the two cortical areas.

Figure 6C samples 12 more of the 37 cognate pair coupling matrices predicted for VISp using Eqn. 1. The 13 representative VISp coupling matrices displayed in **Figs. 6A** and **6C** exemplify the wide variety of neuron-type coupling motifs resulting from transcriptomic prediction (results from ALM are similar but not shown). Most matrices, such as those for pairs *1, 7, 10, 11, 13, 19, 21, 26* and *31*, predict significant coupling over wide swaths of type-pairs, approaching 20% of the entire matrix for pairs *7* and *11*. At the other extreme, a few matrices, such as *29*, 32 and *33*, predict positive coupling between just a very few type pairs (necessitating the circular "magnifying glass" insets in Figure 6C). Other coupling predictions, such as those for pairs *18* and *31* are intermediate in their extent of full matrix coverage.

Figure 6 also illustrates the tendency of positive predictions from individual cognate NP pairs to fall in contiguous "patches" of the full coupling matrix. This is a natural reflection of the strong tendency of both NPP and NP-GPCR expression to align with early nodes in the 2018 Tasic hierarchical clustering which was also evident in **Figs. 3** and **5.** The exemplar matrices of **Figs. 6A** and **6B** include Vip and Vipr1 expression profiles strips along left and top margins to illustrate how patches of coupling within the *1:Vip→Vipr1* matrix reflect Vip and Vipr1 expression profiles. The broadest example of coupling matrix patches reflecting hierarchical taxonomy structure is provided by the observation that large coupling patches fall strictly within single quadrants of glutamatergic-GABAergic signaling as defined by the dashed white crosses in **Fig. 6**.

Differential neuropeptide connectivity predictions from 37 cognate NP pairs densely tile cortical neuron-type coupling matrices. **Figure 7** is a unified visualization of potential neuromodulatory impacts of 37 distinct predicted peptidergic networks, using primary Gα I/S/Q class (see **Table 2**) to approximate signal transduction impacts of each of the 29 NP-GPCRs. As discussed in **Introduction**, it is now known that Gα triage substantially oversimplifies GPCR signal transduction, but such triage should nonetheless provide a worthwhile first approximation. In **Fig. 7**, I, S and Q transduction classes are represented by red, green and blue color channels, respectively, allowing aggregation of all 37 coupling matrices (each individually computed using Eqn. 1, as for **Fig. 6**) according to each NP-GPCR's primary Gα type. The dashed white crosses on both plots in **Fig. 7** divide these connectivity matrices into *E-I* quadrants, exactly analogous to the crosses in **Fig. 6.** Major features of these matrices are clearly quite similar for VISp and ALM.

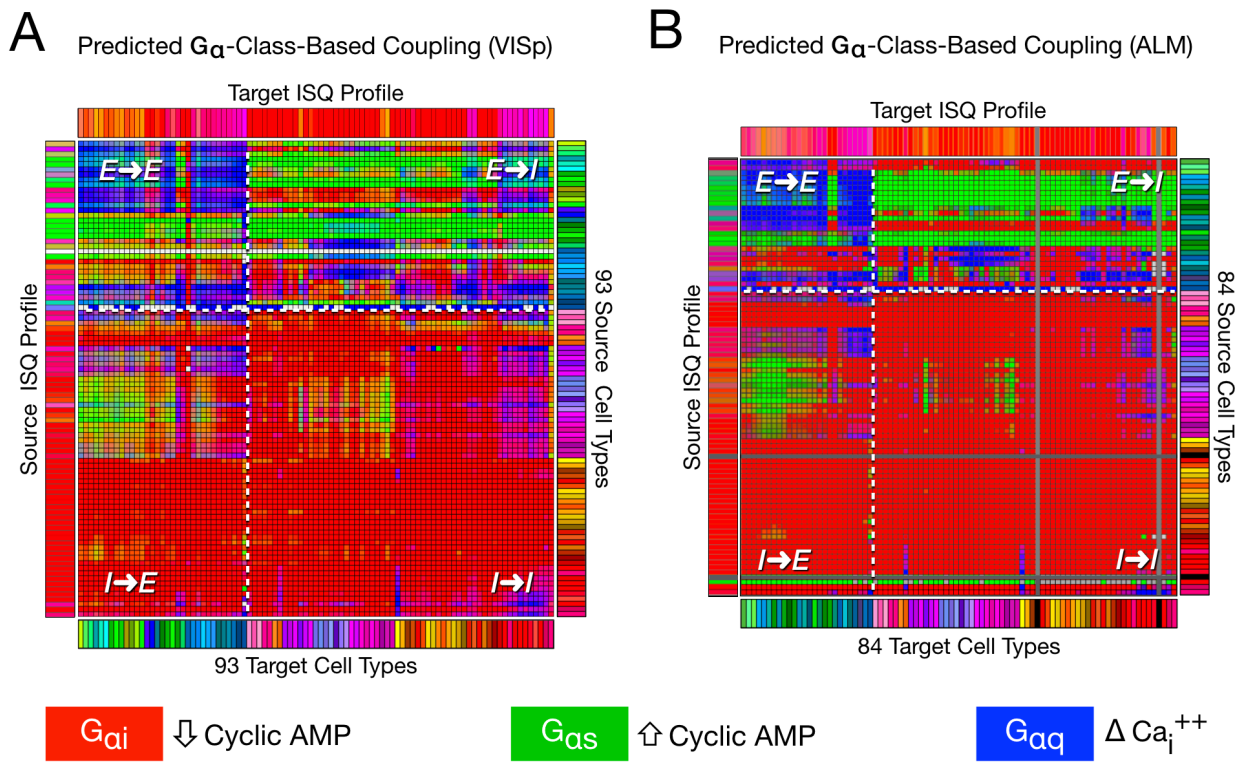


Figure 7. Modulatory coupling predictions based on neuro-type-specific expression of 37 cognate NP gene pairs and primary second-messenger impacts densely tile entire neuron-type local connectivity matrices. Matrix color maps represent predictions of effects of 37 peptidergic coupling networks aggregated, as discussed in main text, on the basis of Gα subtype transduction classes, color-coded in red for class I, green for class S and blue for class Q, as indicated in figure. Axis-label color strips indicate transduction-class profiles summed across target cell types (left axis) and source cell types (top axis), and cell-types (target types at bottom, source types at right). The dashed white cross on each plot demarcates for E-I quadrants analogous to the crosses of Fig. 7: (*E→E*), glutamatergic-to-glutamatergic neuron coupling; (*E→I*), glutamatergic-to-GABAergic coupling; (*I→I*), (GABAergic-to-GABAergic coupling; (*I→E*), GABAergic-to-glutamatergic coupling. **(A)** Predictions of modulatory coupling in area VISp. **(B)** Predictions of modulatory coupling in area ALM. Vertical black bars in type-code color strip and gray bars in heat map here are alignment-preserving spacers marking the two GABAergic cell types absent in ALM. Glutamatergic type alignments differ more substantially (see Tasic, et al. [32] and Supplementary Fig. 3).

The connectivity matrices of **Fig. 7** exhibit three striking features: (1) This aggregate of 37 simplex cognate pair coupling predictions densely tiles the entire neuron-type coupling matrix, (2) the four *E-I* quadrants neatly circumscribe four large rectangular patches dominated by one or the other Gα primary transduction type, Q (blue) in the *E→E* quadrant, S (green) in the *E→I* quadrant, and I (red) in the *I→I* and *I→E* quadrants, and (3) the domains of predicted Gas, Gai and Gaq signal transduction are largely non-overlapping, as evidenced by the rarity of blended colors such as yellow, cyan and magenta. These unexpected features of transduction-class partitioning of the predicted peptidergic coupling matrix suggest that transduction class may have consequences for cortical function that merit further experimental and theoretical investigation.

Discussion

Light from single-cell transcriptomes is now beginning to illuminate dark corners of cellular neuroscience that have long resisted mechanistic and functional analysis [32,43,47–54]. Cortical neuropeptide signaling may be one such corner. While profound impacts of neuropeptide signaling are well-established in a wide range of non-mammalian and sub-cortical neural structures [11,20,22,55–57] and there is a certain amount of excellent literature on cortical NP signaling [58–64], published physiological results are surprisingly rare given the breadth of neuroscientific interest in cortex. The new transcriptomic data analyzed here suggest a possible explanation for this surprising rarity. Though many NPP and many cognate NP-GPCR genes are expressed abundantly in all or very nearly all neocortical neurons, such

expression is highly differential, highly cell-type specific, and highly redundant. Because this multiplicity, cell-type specificity and redundancy have remained uncharted until now, these two factors may have conspired to pose obstacles to repeatable experimentation that may explain the present shortage of published results on cortical neuropeptide mechanisms. The analysis presented above supports this vexatious proposition but may also point the way to productive new perspectives on the roles of local peptidergic neuromodulation in neocortical synaptic network homeostasis and plasticity.

Summary of findings. The present focused analysis of the Tasic 2018 single-cell RNA-Seq dataset establishes that over 97% of mouse neocortical neurons contain mRNA transcripts from one or more NPP genes in high copy numbers and that over 98% contain abundant NP-GPCR transcripts. This observation supports the proposition that all, or very nearly all, neocortical neurons are both neuropeptidergic and modulated by neuropeptides. We are aware of no previous empirical support for such a conclusion. We have closely examined single-neuron expression patterns of sets of 47 NP genes (18 NPP and cognate 29 NP-GPCR) and find that these patterns are very highly conserved between two distant and generally very different areas of neocortex. This observation suggests that NP gene expression must have a very fundamental functional importance. Transcripts of at least one of 18 select NPP genes are detectible in almost every single cortical neuron at extraordinarily high copy number, strongly suggesting very active translation into neuropeptide precursor proteins. Brisk synthesis of precursor proteins establishes, in turn, that processing to active neuropeptides products and brisk secretion of these products is highly probable. We also find that this small set of 47 genes permits transcriptomic neuron-type classification that is exceptionally precise in comparison to otherwise comparable small gene sets, suggesting a very fundamental connection between NP gene expression and neuron type identity. Finally, we have used the Tasic 2018 data to predict architectures of cortical neuropeptidergic signaling networks. There are numerous unexpected and intriguing features of these predictions, including predictions of a very high density of direct coupling between neuron type pairs and of close correlation between neuropeptidergic signal transduction modes and forms of fast synaptic transmission.

Prediction of cortical peptidergic modulation networks. Based on deep, single-cell transcriptomic data, the present analysis predicts the existence of previously unrecognized dense networks of peptidergic neuromodulatory signaling in mouse neocortex. Though these predictions are completely consistent with decades of pioneering work on peptidergic neuromodulation and cortical gene expression [11–13], they leverage the single-cell resolution and genomic depth of the 2018 Tasic data and neuron-type taxonomy to develop many novel predictions and a novel cortical local circuit perspective.

Transcriptomic prediction of local signaling from GABAergic neuron sources is particularly compelling. Because few cortical GABAergic neurons have axons that project beyond the confines of a single cortical area, considerations of diffusion physics and the limited lifetime of peptides after secretion strongly imply that a secreted neuropeptide must act locally, if it acts at all. The high abundance of metabolically expensive NPP transcripts, in turn, offers a strong argument that secreted NPP gene products must have major functional importance. Finally, high cortical expression of NP-GPCR genes cognate to the cortically NPP genes completes an obvious scenario for local peptidergic neuromodulation. These transcriptomic predictions would therefore seem to merit much future experimental and theoretical attention.

Cortical glutamatergic neurons also express NPP genes at high levels that argue for the functional significance of secreted products. Prediction of local neuropeptidergic signaling from cortical glutamatergic neurons nonetheless faces an element of uncertainty that does not confront prediction regarding GABAergic neuron sources. This is because most cortical glutamatergic neurons emit long axons that project far beyond the cell body's local area, so it is possible that all actions of neuropeptides they secrete take place in some remote brain projection target area. While such remote actions are indeed likely, local signaling from glutamatergic neuron sources remains a strong possibility. Most cortical glutamatergic neurons have locally ramifying axons and local presynaptic terminals, and neuropeptides secretion, in any case, is not restricted to axons but can also occur at dendritic and somatic sites. Again, high cortical expression of NP-GPCRs cognate to NPP genes expressed by glutamatergic neurons in the

same local area completes an obvious scenario supportive of local modulatory signaling from glutamatergic neuron sources.

Figures 5, 6 and 7 bifurcate our transcriptome analyses and predictions to focus separately on area VISp and on area ALM, on the premise that paracrine peptidergic signaling *within* the entire extent of a single compact area is quite plausible biophysically, while paracrine signaling *between* these two distant areas is far less plausible. In keeping, however, with the strong conservation of cell-type-specific gene expression profiles between VISp and ALM, the overall forms of predictable peptidergic signaling within the two areas are highly similar. NPP expression is stronger and more varied among the GABAergic neurons of both areas, while NP-GPCR expression is stronger and more varied among the glutamatergic neurons of both areas. This contrast between GABAergic vs. glutamatergic gene expression profiles hints that in both VISp and ALM there is a "prevailing wind" of paracrine peptidergic signaling from GABAergic neurons to glutamatergic neurons. In the present analysis (i.e., in connections with Figs 1BC, 2BC, 5, 6 and 7), we have noted many details of NP gene expression patterning that are strongly conserved between the very different VISp and ALM neocortical areas. We'll argue here, once more, that such conservation supports the proposition that NP gene expression patterns exert major, conserved functional impacts in both VISp and ALM, and by extension perhaps in all cortical areas.

Caveats to functional prediction from transcriptomes. The present predictions of functional neuromodulatory coupling are based on analysis of cellular mRNA abundance, but prediction from such data depends upon (1) extrapolation from cellular mRNA census to inference about the synthesis, processing, localization and functional status of cellular NPP and NP-GPCR proteins, (2) assumptions about neuropeptide diffusion and lifetime in cortical interstitial spaces, (3) assumptions about signal transduction and effector consequences of neuropeptide binding to target cell NP-GPCR receptors. We stipulate that such uncertainties are still substantial and in need of much further investigation, but much is already known and our knowledge is growing rapidly with each advance in single-cell molecular and biophysical methods. Remaining uncertainties notwithstanding, we'll argue here, once more, that the extraordinarily high levels of NPP transcripts detected in most cortical neurons leave little room for reasonable doubt that cortical neurons make use of these transcripts, and the generation and secretion of neuropeptides is the only known potential use. Here we have tried to put forth highly specific predictions about the nature of such usage in the hopes of stimulating research that will find these predictions either prescient or faulty.

Testing peptidergic signaling and network predictions. For predictions of peptidergic signaling and neuromodulatory networks to be of much value, they must be tested anatomically and physiologically. To make the prediction more precise and therefore still more testable, peptide diffusion-reaction dynamics must be characterized biophysically. The analysis above suggests that that experimental attempts at such work in cortex, up until now, would probably have produced few repeatable results due to the uncharted multiplicity, neuron-type-specificity, and redundancy of NPP and NP-GPCR expression characteristic of neocortex. This dismal situation should now improve, however, with the emergence of neuron-type taxonomies and tools for experimental access to specific cortical neuron types. Such access may be either *prospective*, using Cre driver lines [29,65,66] or viral vectors [67] of substantial neuron-type specificity, or *retrospective* using highly multiplexed FISH [36,68] or immunostaining methods [65,69], patch-seq [37,68] or morphological neuron-type classification methods [36,70]. By eliminating confusion amongst myriad cortical neuron types expressing myriad diverse NP signaling genes, these new abilities to access specific cortical neuron types should enormously improve the prospects for repeatable experimental tests of neuron-type-specific peptidergic signaling predictions.

A vast present pharmacopeia of well-characterized specific ligands and antagonists for most NP-GPCRs [46] will be the bedrock for the functional analysis of neuron-type specific peptide signaling. For analysis of type-specific neuropeptide signaling in network context (i.e., *ex vivo* slices and *in vivo*), newer optophysiological methods of calcium imaging and optogenetic stimulation/inhibition will certainly join electrophysiology as the foundation for measurement of neuropeptide impacts. In addition, however, many new tools more specific to neuropeptide signaling are emerging. Super-resolution 3D immunohistologies like array tomography [71] and 3D single-molecule methods [72,73] will enable

imaging of DCV localization and neuropeptide contents in type-specific network anatomical context. Genetically encoded sensors of extracellular GPCR ligands [74], GPCR activation [75–79], G-protein mobilization [79], cAMP concentration [80], protein kinase activation [[81] and protein phosphorylation [77] will enable fine dissection of NP dynamics and NP-GPCR signal transduction events[82]. All of these tools have been demonstrated in principle in physiological applications, and all should be readily applicable to wide ranges of specific neuropeptides and specific neuron types. In addition, new caged NP-GPCR ligands [83] and antagonists [84] will provide new dimensions of spatial and temporal control for NP receptor activation. Combination of this impressive array of old and new tools with new means of experimental access to specific neuron types seems almost certain to enable critical tests of the peptidergic network signaling hypotheses we have hope to inspire with the type-specific predictions we have set forth here.

Theoretical neuroscience prospects. Rapidly growing compute power and advancing theoretical and computational modeling approaches offer the greatest promise at present for fathoming the otherwise daunting complexities of synaptic network function[85–89]. In parallel with advances in computational neuroscience, the computer science of artificial intelligence has increasingly embraced artificial neural net (ANN) architectures inspired historically by mid-twentieth century views of biological neuronal network architectures [90–94]. Interestingly, one of the hardest problems facing both neuroscience and computer science threads is shared between the disciplines: the adaptive adjustment, one by one, of very large numbers of what both fields call “synaptic weights”. For useful function of both biological and artificial networks, large numbers of synaptic weights must be somehow dynamically adjusted during developmental and “learning” (or “training”) processes. The core hard problem is the assignment of “credit” for success or failure to the right individual synapses out of the very many, in order to guide subsequent individualized adjustment. Theoretical neuroscientists struggle with this “credit assignment” problem over the discovery and empirical validation of biologically plausible learning rules. Computer scientists struggle with the extreme computational expense of currently standard backpropagation-based ANN re-weighting algorithms. This re-convergence of biological and computer science interests has accordingly driven very productive efforts lately at the intersection of disciplines. These have included both biological mapping of synaptic and molecular connectomes in ever-increasing detail, and computer scientists’ interest in another round of biological inspiration, aimed this time aimed at learning how evolution may have finessed the hard credit assignment problem [95–97].

The science of neuromodulation, as discussed briefly in our **Introduction,** addresses directly the mechanisms by while synaptic weights and firing properties are adjusted in biological synaptic networks. It therefore seems possible that new transcriptomic insights into neuromodulatory signaling may address the shared desires of computational neuroscientists and computer scientists for a deeper understanding of the biological principles of credit assignment and synaptic weight adjustment. As one case in point, the broad class of synaptic learning rules that neuroscientists call “spike-timing-dependent plasticity” (STDP) [98,99] has already attracted a great deal of attention by computer scientists as a possible element of more efficient credit assignment and ANN parameter adjustment algorithms, particularly when STDP is gated by modulatory signals [4,100–106].

While most biological studies of STDP gating to date have referenced the monoamine neuromodulators [100,107,108], known commonalities of signal transduction downstream from widely varying GPCRs suggest strongly that NP-GPCRs play roles closely analogous to those postulated for dopamine-selective GPCRs [106]. Furthermore, the much greater neuron-type-specificity of NP-GPCRs may overcome the shortcomings of the purely monoamine-based STDP gating that have emerged in computational studies to date. With ongoing progress in the development of cell-type-based synaptic connectomes and cell-type-based modulation networks, the day may soon come for a neuroscience modeling of synaptic networks in which empirically-informed synaptic parameters are adjusted via empirically-informed neuromodulatory signaling. The present predictions of peptidergic modulation networks may bring that day a little closer.

Prospects for psychiatric drug development. Comprehensive transcriptomic views of cortical neuropeptide signaling may be poised to open a new era for molecular psychiatry and the development

of novel GPCR-based psychiatric pharmaceuticals. Recent decades of molecular psychiatry serve as proof of principle that treatments aimed at manipulating neuromodulatory GPCR signaling can profoundly alter the progression of psychiatric disorders such as depression, autism and schizophrenia, but also revealed many serious limitations of the present pharmacopoeia. The psychiatric pharmaceuticals in question almost all target signaling involving the monoamine neuromodulators dopamine, serotonin, noradrenaline and/or histamine and their selective GPCR receptors [109–112]. While there have been many efforts to develop psychiatric pharmaceuticals targeting neuropeptide-based GPCR signaling [18,113], successes to date have been relatively limited, outside the realm of the very heavily and successfully "drugged" opioid neuropeptide receptors.

The present study raises the possibility that the same uncharted multiplicity, cell-type-specificity and redundancy of NPP and NP-GPCR expression that may have hampered the physiological study of cortical neuropeptide signaling may have also hindered the development of a neuropeptide-based molecular psychiatry. Here we propose that this obstacle might be transformed into a garden path by new transcriptomic maps of NPP and NP-GPCR cell-type specificity. The greater neuron-type and neuroanatomical specificity of neuropeptide signaling as compared to monoamine signaling suggests the very attractive possibility of reducing side-effects by targeting the more neuron-type-specific neuropeptide systems. Moreover, while GPCRs have long been known as among the most "druggable" of targets [114,115], the druggability of GPCRs is currently advancing very rapidly due to advances in GPCR structural biology and molecular dynamic simulations [116–118]. Already the basic science of neuropeptide biology has benefitted greatly from advances in genomic, transcriptomic and structural biology methods and it seems inevitable these advances will translate into advances addressing the many neurological and psychiatric disorders may have roots in dysregulation of cortical synaptic networks.

Acknowledgements

We wish to thank the Allen Institute for Brain Science founder, Paul G. Allen, for his vision, encouragement and support. Many other colleagues within and beyond the Allen Institute (to be named in next revision) have contributed to development of the ideas we have put forth here. This work was supported in part by award number R01NS092474 from the Office of the Director of National Institutes of Health and award number R01MH104227 from the National Institute of Mental Health. The content is solely the responsibility of the authors and does not necessarily represent official views of the National Institutes of Health.

Bibliography

1. Mccormick DA, Nusbaum MP. Editorial overview: Neuromodulation: Tuning the properties of neurons, networks and behavior. Curr Opin Neurobiol. 2014;29:4–7.

2. Marder E, Goeritz ML, Otopalik AG. Robust circuit rhythms in small circuits arise from variable circuit components and mechanisms. Curr Opin Neurobiol. 2015;31:156–63.

3. Nadim F, Bucher D. Neuromodulation of neurons and synapses. Curr Opin Neurobiol. 2014;29:48–56.

4. Pawlak V, Wickens JR, Kirkwood A, Kerr JND. Timing is not everything: Neuromodulation opens the STDP gate. Front Synaptic Neurosci. 2010;2:1–14.

5. Abbott LF, Regehr WG. Synaptic computation. Nature. 2004;431:796–803.

6. Bargmann CI. Beyond the connectome: How neuromodulators shape neural circuits. BioEssays. 2012;34:458–65.

7. Nusbaum MP, Blitz DM, Marder E. Functional consequences of neuropeptide and small-molecule co-transmission. Nat Rev Neurosci. 2017;18:389–403.

8. Bucher D, Marder E. XSnapShot: Neuromodulation. Cell [Internet] 2013;155:482–482.e1.

9. Swinehart CD, Abbott LF. Supervised learning through neuronal response modulation. Neural Comput. 2005;17:609–31.

10. Marder E. Neuromodulation of Neuronal Circuits: Back to the Future. Neuron. 2012;76:1–11.

11. Burbach JPH. What are neuropeptides? Merighi A, editor. Methods Mol Biol. 2011;789:1–36.

12. van den Pol AN. Neuropeptide Transmission in Brain Circuits. Neuron. 2012;76:98–115.

13. Hökfelt T, Ögren SO, Xu ZQD. Classical Neurotransmitters and Neuropeptides. Handb. Biol. Act. Pept., Second Edition. 2013; 1835-1842

14. Wang Y, Wang M, Yin S, Jang R, Wang J, Xue Z, Xu T. NeuroPep: A comprehensive resource of neuropeptides. Database. 2015;2015:1–9.

15. Gonzalez-Suarez AD, Nitabach MN. Peptide-Mediated Neurotransmission Takes Center Stage. Trends Neurosci. 2018;41:325–7.

16. Russo AF. Overview of Neuropeptides: Awakening the Senses? Headache. 2017;57:37–46.

17. Krishnan A, Schioth HB. The role of G protein-coupled receptors in the early evolution of neurotransmission and the nervous system. J Exp Biol. 2015;218:562–71.

18. Hoyer D, Bartfai T. Neuropeptides and neuropeptide receptors: Drug targets, and peptide and non-peptide ligands: A tribute to prof. dieter seebach. Chem Biodivers. 2012;9:2367–87.

19. Varoqueaux F, Fasshauer D. Getting Nervous: An Evolutionary Overhaul for Communication. Annu Rev Genet. 2017;51:455–76.

20. Elphick MR, Mirabeau O, Larhammar D. Evolution of neuropeptide signalling systems. J Exp Biol. 2018;221:jeb151092.

21. Jekely G. Global view of the evolution and diversity of metazoan neuropeptide signaling. Proc Natl Acad Sci. 2013;110:8702–7.

22. Grimmelikhuijzen CJP, Hauser F. Mini-review: The evolution of neuropeptide signaling. Regul Pept. 2012;177:S6–9.

23. Pignatelli M, Bonci A. Role of Dopamine Neurons in Reward and Aversion: A Synaptic Plasticity Perspective. Neuron. 2015;86:1145–57.

24. Markram H, Gerstner W, Sjöström PJ. Spike Timing-Dependent Plasticity: A Comprehensive Overview. Front. Synaptic Neurosci. 2013. pp. 1-574.

25. Ludwig M, Leng G. Dendritic peptide release and peptide-dependent behaviours. Nat. Rev. Neurosci. 2006. p. 126–36.

26. Nässel DR. Neuropeptide signaling near and far: How localized and timed is the action of neuropeptides in brain circuits? Invertebr Neurosci. 2009;9:57–75.

27. Tremblay R, Lee S, Rudy B. GABAergic Interneurons in the Neocortex: From Cellular Properties to Circuits. Neuron. 2016;91:260–92.

28. Maximiliano José N, Hashikawa Y, Rudy B. Diversity and connectivity of layer 5 somatostatin-expressing interneurons in the mouse barrel cortex. J Neurosci . 2018;38:2415–7.

29. Daigle TL, Madisen L, Hage TA, Valley MT, Knoblich U, Larsen RS, Takeno MM, Huang L, Gu H, Larsen R, Mills M, Bosma-Moody A, Siverts LA, Walker M, Graybuck LT, Yao Z, et al. A Suite of Transgenic Driver and Reporter Mouse Lines with Enhanced Brain-Cell-Type Targeting and Functionality. Cell. 2018;174:465–480.e22.

30. Lein E, Hawrylycz MJ, Ao N, Ayres M, Bensinger A, Bernard A, Boe AF, Boguski MS, Brockway KS, Byrnes EJ, Chen L, Chen L, Chen TM, Chin MC, Chong J, Crook BE, et al. Genome-wide atlas of gene expression in the adult mouse brain. Nature . 2007;445:168–76.

31. Hawrylycz M, Lein E, Guillozet-Bongaarts AL, Shen EH, Ng L, Miller JA, Van De Lagemaat LN, Smith KA, Ebbert A, Riley ZL, Abajian C, Beckmann CF, Bernard A, Bertagnolli D, Boe AF, Cartagena PM, et al. An anatomically comprehensive atlas of the adult human brain transcriptome. Nature. 2012;489:391–9.

32. Tasic B, Yao Z, Smith KA, Graybuck L, Nguyen TN, Bertagnolli D, Goldy J, Garren E, Economo MN, Viswanathan S, Penn O, Bakken T, Menon V, Miller JA, Fong O, Hirokawa KE, et al. Shared and distinct transcriptomic cell types across neocortical areas. Nature. 2018;563:72–8.

33. Wang X, Allen WE, Wright MA, Sylwestrak EL, Samusik N, Vesuna S, Evans K, Liu C, Ramakrishnan C, Liu J, Nolan GP, Bava FA, Deisseroth K. Three-dimensional intact-tissue sequencing of single-cell transcriptional states. Science. 2018;361:1-9.

34. Moffitt JR, Hao J, Bambah-mukku D, Lu T, Dulac C, Zhuang X. High-performance multiplexed fluorescence in situ hybridization in culture and tissue with matrix imprinting and clearing. Proc Natl Acad Sci . 2016;113:201617699.

35. Shah S, Lubeck E, Schwarzkopf M, He T-F, Greenbaum A, Sohn CH, Lignell A, Choi HMT, Gradinaru V, Pierce NA, Cai L. Single-molecule RNA detection at depth by hybridization chain reaction and tissue hydrogel embedding and clearing. Development. 2016;143:2862–7.

36. Zeng H, Sanes JR. Neuronal cell-type classification: Challenges, opportunities and the path forward. Nat Rev Neurosci. 2017;18:530–46.

37. Cadwell CR, Sandberg R, Jiang X, Tolias AS. Q&A: Using Patch-seq to profile single cells. BMC Biol. BMC Biology. 2017;15:1–7.

38. Tasic B. Single cell transcriptomics in neuroscience: cell classification and beyond. Curr Opin Neurobiol. 2018;50:242–9.

39. Allen Brain Map: Cell Taxonomies [Internet]. https://portal.brain-map.org/explore/classes. [cited 2019 Jan 7].

40. Allen Brain Atlas: Data Portal [Internet]. http://celltypes.brain-map.org/rnaseq/. [cited 2019 Jan 7].

41. Suter DM, Molina N, Gatfield D, Schneider K, Schibler U, Naef F. Mammalian genes are transcribed with widely different bursting kinetics. Science. 2011;332:472–4.

42. Kim JK, Kolodziejczyk AA, Illicic T, Teichmann SA, Marioni JC. Characterizing noise structure in single-cell RNA-seq distinguishes genuine from technical stochastic allelic expression. Nat Commun. 2015;6.

43. Tasic B, Menon V, Nguyen TNT, Kim TKTTK, Jarsky T, Yao Z, Levi BB, Graybuck L, Sorensen SA, Dolbeare T, Bertagnolli D, Goldy J, Shapovalova N, Parry S, Lee CC, Smith K, et al. Adult mouse cortical cell taxonomy revealed by single cell transcriptomics. Nat Neurosci . 2016;19:335–46.

44. Hastie T, Tibshirani R, Friedman J. The Elements of Statistical Learning: Data Mining, Inference, Prediction (pages 369-370). Springer US; 2001.

45. Oyelade J, Isewon I, Oladipupo F, Aromolaran O, Uwoghiren E, Aameh F, Aachas M, Aadebiyi E. Clustering algorithms: Their application to gene expression data. Bioinform Biol Insights. 2016;10:237–53.

46. IUPHAR/BPS. GPCR Neuropharmacology Guide [Internet]. http://www.guidetopharmacology.org/GRAC/ReceptorFamiliesForward?type=GPCR. [cited 2019 Jan 6].

47. Telley L, Govindan S, Prados J, Stevant I, Nef S, Dermitzakis E, Dayer A, Jabaudon D. Sequential transcriptional waves direct the differentiation of newborn neurons in the mouse neocortex. Science. 2016;351:1443–6.

48. Shekhar K, Lapan SW, Whitney IE, Tran NM, Macosko EZ, Kowalczyk M, Adiconis X, Levin JZ, Nemesh J, Goldman M, McCarroll SA, Cepko CL, Regev A, Sanes JR. Comprehensive Classification of Retinal Bipolar Neurons by Single-Cell Transcriptomics. Cell. 2016;166:1308–1323.e30.

49. Okaty BW, Sugino K, Nelson SB. Cell Type-Specific Transcriptomics in the Brain. J Neurosci . 2011;31:6939–43.

50. Fan X, Dong J, Zhong S, Wei Y, Wu Q, Yan L, Yong J, Sun L, Wang X, Zhao Y, Wang W, Yan J, Wang X, Qiao J, Tang F. Spatial transcriptomic survey of human embryonic cerebral cortex by single-cell RNA-seq analysis. Cell Res. 2018;28:730–45.

51. Paul A, Crow M, Raudales R, He M, Gillis J, Huang ZJ. Transcriptional Architecture of Synaptic Communication Delineates GABAergic Neuron Identity. Cell. 2017;171:522–539.e20.

52. Zeng H, Sanes JR. Neuronal cell-type classification: challenges, opportunities and the path forward. Nat Rev Neurosci. 2017;18:530-546.

53. Földy C, Darmanis S, Aoto J, Malenka RC, Quake SR, Südhof TC. Single-cell RNAseq reveals cell adhesion molecule profiles in electrophysiologically defined neurons. Proc Natl Acad Sci . 2016;113:E5222–31.

54. Gokce O, Stanley GM, Treutlein B, Neff NF, Camp JG, Malenka RC, Rothwell PE, Fuccillo M V., Südhof TC, Quake SR. Cellular Taxonomy of the Mouse Striatum as Revealed by Single-Cell RNA-Seq. Cell Rep. 2016;16:1126–37.

55. Kuffler SW, Jan LY, Jan YN. A peptide as a possible transmitter in sympathetic ganglia of the frog. PNAS. 1979;76:1501–5.

56. Katz PS, Lillvis JL. Reconciling the deep homology of neuromodulation with the evolution of behavior. Curr Opin Neurobiol. 2014;29:39–47.

57. Borbély É, Scheich B, Helyes Z. Neuropeptides in learning and memory. Neuropeptides. 2013;47:439–50.

58. Williams JT, Zieglgänsberger W. Neurons in the frontal cortex of the rat carry multiple opiate receptors. Brain Res. 1981;226:304–8.

59. Rossier J, Chapouthier G. Brain opiates. Endeavour. 1982;6:168–76.

60. Gallopin T, Geoffroy H, Rossier J, Lambolez B. Cortical sources of CRF, NKB, and CCK and their effects on pyramidal cells in the neocortex. Cereb Cortex. 2006;16:1440–52.

61. Gomtsian L, Bannister K, Eyde N, Robles D, Dickenson AH, Porreca F, Navratilova E. Morphine effects within the rodent anterior cingulate cortex and rostral ventromedial medulla reveal separable modulation of affective and sensory qualities of acute or chronic pain. Pain . 2018;159:1.

62. Liu JJ, Sharma K, Zangrandi L, Chen C, Humphrey SJ, Chiu YT, Spetea M, Liu-Chen LY, Schwarzer C, Mann M. In vivo brain GPCR signaling elucidated by phosphoproteomics. Science. 2018;360:1-10.

63. Férézou I, Hill EL, Cauli B, Gibelin N, Kaneko T, Rossier J, Lambolez B. Extensive overlap of mu-opioid and nicotinic sensitivity in cortical interneurons. Cereb Cortex. 2007;17:1948–57.

64. Mena JD, Selleck RA, Baldo BA. Mu-Opioid Stimulation in Rat Prefrontal Cortex Engages Hypothalamic Orexin/Hypocretin-Containing Neurons, and Reveals Dissociable Roles of Nucleus Accumbens and Hypothalamus in Cortically Driven Feeding. J Neurosci . 2013;33:18540–52.

65. He M, Tucciarone J, Lee S, Nigro MJJ, Kim Y, Levine JMM, Kelly SMM, Krugikov I, Wu P, Chen Y, Gong L, Hou Y, Osten P, Rudy B, Huang ZJ. Strategies and Tools for Combinatorial Targeting of GABAergic Neurons in Mouse Cerebral Cortex. Neuron . 2016;92:555.

66. Madisen L, Garner ARR, Shimaoka D, Chuong ASS, Klapoetke NCC, Li L, van der Bourg A, Niino Y, Egolf L, Monetti C, Gu H, Mills M, Cheng A, Tasic B, Nguyen TNN, Sunkin SMM, et al. Transgenic mice for intersectional targeting of neural sensors and effectors with high specificity and performance. Neuron . 2015;85:942–58.

67. Dimidschstein J, Chen Q, Tremblay R, Rogers SL, Saldi GA, Guo L, Xu Q, Liu R, Lu C, Chu J, Grimley JS, Krostag AR, Kaykas A, Avery MC, Rashid MS, Baek M, et al. A viral strategy for targeting and manipulating interneurons across vertebrate species. Nat Neurosci . 2016;19:1743–9.

68. Lein E, Borm LE, Linnarsson S. The promise of spatial transcriptomics for neuroscience in the era of molecular cell typing. Science. 2017;358:64–9.

69. Xu X, Roby KD, Callaway EM. Immunochemical characterization of inhibitory mouse cortical neurons: Three chemically distinct classes of inhibitory cells. J Comp Neurol. 2010;518:389–404.

70. DeFelipe J, López-Cruz PL, Benavides-Piccione R, Bielza C, Larrañaga P, Anderson S, Burkhalter A, Cauli B, Fairén A, Feldmeyer D, Fishell G, Fitzpatrick D, Freund TF, González-Burgos G, Hestrin S, Hill S, et al. New insights into the classification and nomenclature of cortical GABAergic interneurons. Nat Rev Neurosci. 2013;14:202–16.

71. Smith SJ. Q&A: Array tomography. BMC Biology; 2018;16:1–18.

72. Jia S, Vaughan JC, Zhuang X. Isotropic three-dimensional super-resolution imaging with a self-bending point spread function. Nat Photonics. 2014;8:302–6.

73. von Diezmann A, Shechtman Y, Moerner WE. Three-Dimensional Localization of Single Molecules for Super-Resolution Imaging and Single-Particle Tracking. Chem Rev . 2017;acs.chemrev.6b00629.

74. Patriarchi T, Cho JR, Merten K, Howe MW, Marley A, Xiong WH, Folk RW, Broussard GJ, Liang R, Jang MJ, Zhong H, Dombeck D, von Zastrow M, Nimmerjahn A, Gradinaru V, Williams JT, et al. Ultrafast neuronal imaging of dopamine dynamics with designed genetically encoded sensors. Science. 2018;360. doi: 10.1126/science.aat4422. Epub 2018 May 31.

75. Kauk M, Hoffmann C. Intramolecular and Intermolecular FRET Sensors for GPCRs - Monitoring Conformational Changes and Beyond. Trends Pharmacol Sci. 2018;39(2):123-135.

76. Stoeber M, Jullié D, Lobingier BT, Laeremans T, Steyaert J, Schiller PW, Manglik A, von Zastrow M. A Genetically Encoded Biosensor Reveals Location Bias of Opioid Drug Action. Neuron. 2018;963–76.

77. Haider RS, Godbole A, Hoffmann C. To sense or not to sense—new insights from GPCR-based and arrestin-based biosensors. Curr Opin Cell Biol. 2019;57:16–24.

78. Livingston KE, Mahoney JP, Manglik A, Sunahara RK, Traynor JR. Measuring ligand efficacy at the mu-opioid receptor using a conformational biosensor. Elife. 2018;7:1–23.

79. Ratnayake K, Kankanamge D, Senarath K, Siripurapu P, Weis N, Tennakoon M, Payton JL, Karunarathne A. Measurement of GPCR-G protein activity in living cells. 2nd ed. Methods Cell Biol. 2017;142:1-25.

80. Hackley CR, Mazzoni EO, Blau J. cAMPr: A single-wavelength fluorescent sensor for cyclic AMP. Sci Signal. 2018;11.

81. Chen Y, Saulnier JL, Yellen G, Sabatini BL. A PKA activity sensor for quantitative analysis of endogenous GPCR signaling via 2-photon FRET-FLIM imaging. Front Pharmacol. 2014;5 APR:1–12.

82. Spangler SM, Bruchas MR. Optogenetic approaches for dissecting neuromodulation and GPCR signaling in neural circuits. Curr Opin Pharmacol. 2017;32:56–70.

83. Banghart MR, He XJ, Sabatini BL. A Caged Enkephalin Optimized for Simultaneously Probing Mu and Delta Opioid Receptors. ACS Chem Neurosci. 2018;9:684–90.

84. Banghart MR, Williams JT, Shah RC, Lavis LD, Sabatini BL. Caged Naloxone Reveals Opioid Signaling Deactivation Kinetics. Mol Pharmacol. 2013;84:687–95.

85. Vreeswijk C Van, Sompolinsky H. Chaos in Neuronal Networks with Balanced Excitatory and Inhibitory Activity. Science. 1996;274:1724–6.

86. Koch C, Segev I. Methods in Neuronal Modeling: From Ions to Networks. Koch C, Segev I, editors. Comput. Neurosci. MIT Press; 1998.

87. Song S, Miller KD, Abbott LF. Competitive Hebbian learning through spike-timing-dependent synaptic plasticity. Nat Neurosci. 2000;3:919–26.

88. Dayan P, Abbott LF. Theoretical Neuroscience - Computational and Mathematical Modeling of Neural Systems. MIT Press; 2001.

89. Hopfield JJ. Neural networks and physical systems with emergent collective computational abilities. Proc Natl Acad Sci U S A. 1982;79:2554–8.

90. Rosenblatt F. The perceptron: A probabilistic model for information storage and organization in the brain. Psychol Rev. 1958;65:386–408.

91. M. Minsky, Papert S. Perceptrons, Expanded Edition An Introduction to Computational Geometry. MIT Press. 1969.

92. Rumelhart DE, McClelland JL, PDP Research Group. Parallel Distributed Processing. Massachusetts Inst. Technol. 1987.

93. Ballard DH, Hinton GE, Sejnowski TJ. Parallel visual computation. Nature. 1983;306:21–6.

94. Lecun Y, Bengio Y, Hinton GE. Deep learning. 2015;

95. Guerguiev J, Lillicrap TP, Richards BA. Towards deep learning with segregated dendrites. Elife. 2017;6:1–2.

96. Lillicrap TP, Cownden D, Tweed DB, Akerman CJ. Random synaptic feedback weights support error backpropagation for deep learning. Nat Commun. 2016;7:1–10.

97. Huh D, Sejnowski TJ. Gradient Descent for Spiking Neural Networks [Internet] arXiv:1706.04698. 2017.

98. Dan Y, Poo M-M. Spike Timing-Dependent Plasticity: From Synapse to Perception. Physiol Rev. 2006;86:1033–48.

99. Poo M-M, Pignatelli M, Ryan TJ, Tonegawa S, Bonhoeffer T, Martin KC, Rudenko A, Tsai LH, Tsien RW, Fishell G, Mullins C, Gonçalves JT, Shtrahman M, Johnston ST, Gage FH, Dan Y, et al. What is memory? The present state of the engram. BMC Biology; 2016;14:1–18.

100. Izhikevich EM. Solving the distal reward problem through linkage of STDP and dopamine signaling. Cereb Cortex. 2007;17:2443–52.

101. Xie X, Seung HS. Equivalence of Backpropagation and Contrastive Hebbian. Neural Comput. 2003;454:441–54.

102. Florian R. Reinforcement learning through modulation of spike-timing dependent synaptic plasticity. Neural Comput. 2007;19:1468–1502.

103. Farries MA, Fairhall AL. Reinforcement Learning With Modulated Spike Timing Dependent Synaptic Plasticity. J Neurophysiol. 2007;98:3648–65.

104. Bengio Y, Lee D-H, Bornschein J, Mesnard T, Lin Z. Towards Biologically Plausible Deep Learning. 2016;

105. Frémaux N, Gerstner W. Neuromodulated Spike-Timing-Dependent Plasticity, and Theory of Three-Factor Learning Rules. Front Neural Circuits. 2016;9.

106. Roelfsema PR, Holtmaat A. Control of synaptic plasticity in deep cortical networks. Nat Rev Neurosci. 2018;19:166–80.

107. Schultz W. Neuronal Reward and Decision Signals: From Theories to Data. Physiol Rev. 2015;95:853–951.

108. Kuśmierz Ł, Isomura T, Toyoizumi T. Learning with three factors: modulating Hebbian plasticity with errors. Curr Opin Neurobiol. 2017;46:170–7.

109. Hamon M, Blier P. Monoamine neurocircuitry in depression and strategies for new treatments. Prog Neuro-Psychopharmacology Biol Psychiatry. 2013;45:54–63.

110. Data-Franco J, Singh A, Popovic D, Ashton M, Berk M, Vieta E, Figueira ML, Dean OM. Beyond the therapeutic shackles of the monoamines: New mechanisms in bipolar disorder biology. Prog Neuro-Psychopharmacology Biol Psychiatry. 2017;72:73–86.

111. Millan MJ, Goodwin GM, Meyer-Lindenberg A, Ove Ogren S. Learning from the past and looking to the future: Emerging perspectives for improving the treatment of psychiatric disorders. Eur Neuropsychopharmacol. 2015;25:599–656.

112. Urs NM, Nicholls PJ, Caron MG. Integrated approaches to understanding antipsychotic drug action at GPCRs. Curr Opin Cell Biol. 2014;27:56–62.

113. Hökfelt T, Bartfai T, Bloom F. Neuropeptides: Opportunities for drug discovery. Lancet Neurol. 2003;2:463–72.

114. Gurrath M. Peptide-Binding G Protein-Coupled Receptors: New Opportunities for Drug Design. Curr Med Chem. 2001;8:1605–48.

115. Lundstrom K. An Overview on GPCRs and Drug Discovery: Structure-Based Drug Design and Structural Biology on GPCRs. In: Leifert WR, editor. G Protein-Coupled Recept Drug Discov. Totowa, NJ: Humana Press; 2009. p. 51–66.

116. Hilger D, Masureel M, Kobilka BK. Structure and dynamics of GPCR signaling complexes. Nat Struct Mol Biol. 2018;25:4–12.

117. Koehl A, Hu H, Maeda S, Zhang Y, Qu Q, Paggi JM, Latorraca NR, Hilger D, Dawson R, Matile H, Schertler GFX, Granier S, Weis WI, Dror RO, Manglik A, Skiniotis G, et al. Structure of the μ-opioid receptor–Gi protein complex. Nature. 2018;558:547–52.

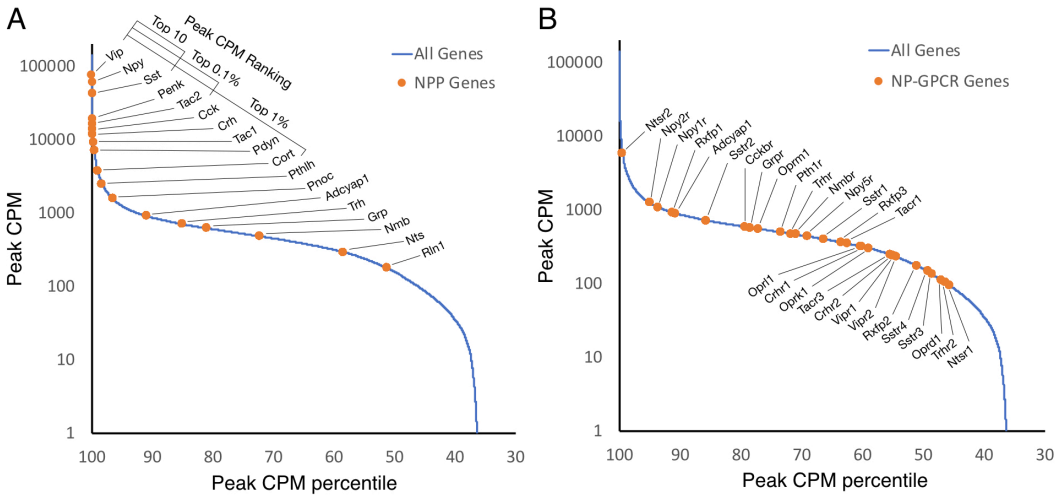
118. Weis WI, Kobilka BK. The Molecular Basis of G Protein-Coupled Receptor Activation. Annu Rev Biochem. 2018;87:897–919.

119. Abadi M, Agarwal A, Barham P, Brevdo E, Chen Z, Citro C, Corrado GS, Davis A, Dean J, Devin M, Ghemawat S, Goodfellow I, Harp A, Irving G, Isard M, Jia Y, et al. TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems. 2016;

120. Choilet F, et al. Keras: The Python Deep Learning Library [Internet]. https://keras.io.

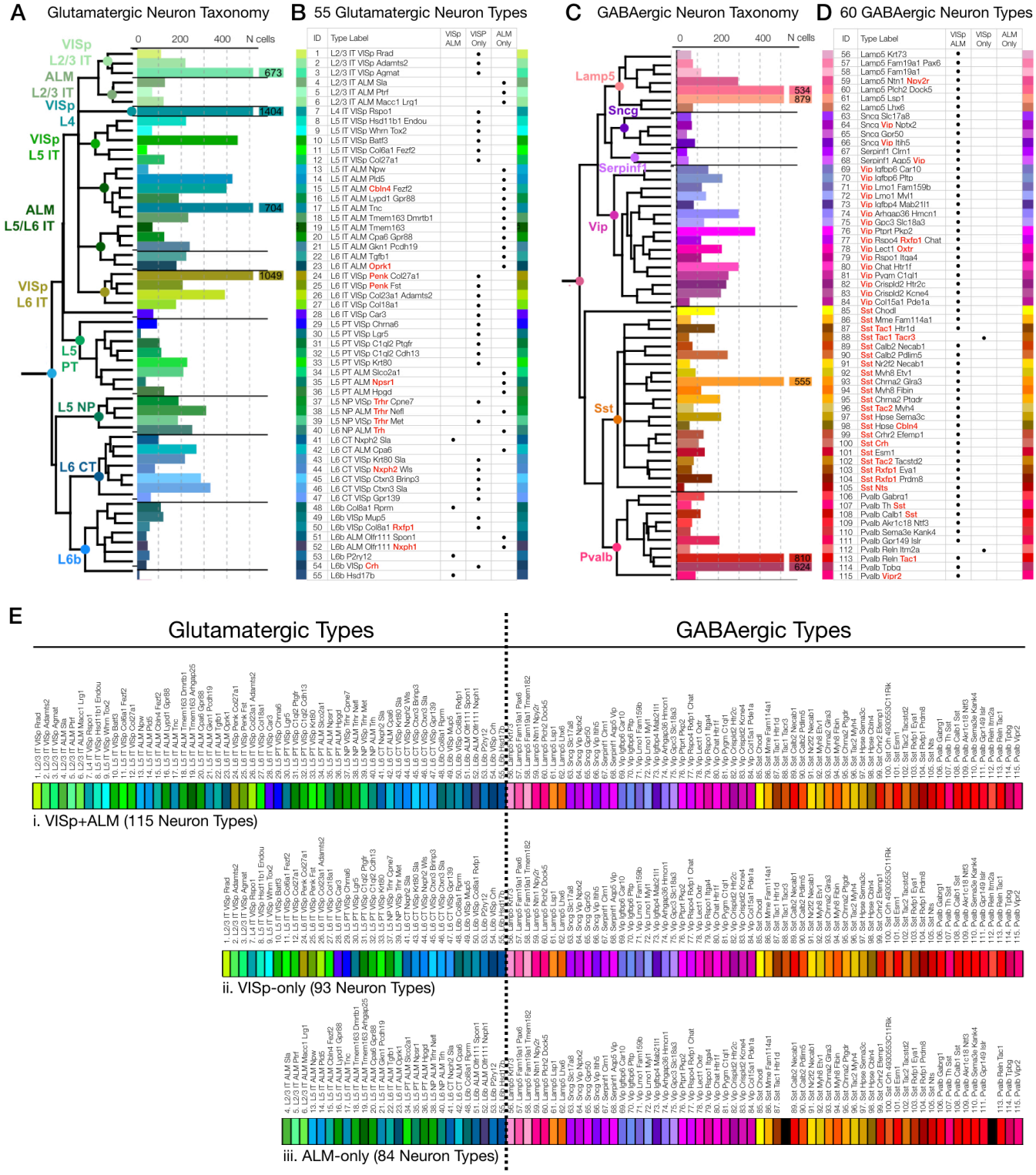
Supplementary Materials

Supplementary Figure 1. Peak CPM values for 18 NPP and 29 NP-GPCR genes in the context of peak CPM values for all cortically expressed protein-coding genes.



Summary of results from single-cell mRNA-Seq analysis of 22,439 individual neurons from cortical areas VISp and ALM. The identical solid curves in (A) and (B) represent ranked peak CPM values (see main text) of all 21,931 protein-coding genes detected across all neurons plotted against percentiles peak CPM rank. (A) Filled circles indicate peak CPM values for the 18 NPP genes of Table 1, as indicated by callouts on graph. Transcripts of these 18 NPP genes are among the very most abundant messages in individual mouse cortical neurons. Slanted brackets highlight three 3 NPP genes ranking in the top 10 for all genes by single-cell peak CPM, five that fall within the top 0.1% and ten within the top 1%. All 18 are in the top 50% by this abundance metric. (B) Filled circles indicate peak CPM vales are for the cognate 29 NP-GPCR genes of Table 1. Almost all of these NP-GPCR genes exhibit single-cell expression levels ranking in the top 50% of those for all protein-coding genes.

Supplementary Figure 2. Brief summary of the 2018 Tasic [32] neuron-type taxonomy.



Transcriptomic neuron-type taxonomies enable neuron-type-specific profiling of NPP and NP-GPCR gene expression. Tasic and co-workers [32] generated these taxonomies by iterative hierarchical clustering of data from single-cell RNA-Seq analysis of 22,439 mouse cortical neurons dissociated from areas VISp and ALM. (A) Taxonomic cladogram of glutamatergic neurons. Horizontal bars represent numbers of analyzed cells falling into each neuron type category. (B) Neuron type labels and type-code

colors for each glutamatergic neuron type. (C) Taxonomic cladogram for GABAergic neurons. (D) Neuron type labels and type-code colors for each GABAergic neuron type. (E) Concatenated glutamatergic and GABAergic type-code color strips and type labels for (i) reference aggregate of VISp and ALM neuron types (VISp+ALM), (ii) VISp-only neuron types, and (iii) ALM-only neuron types. The neuron-type color codes and taxonomy strips i-iii will be re-used consistently throughout the present publication.

The cladogram in (A) represents the Tasic 2018 taxonomy of glutamatergic neurons in areas VISp and ALM and indicates the number of single-cell transcriptomes mapped to each “leaf” cluster. These cell numbers bear some relation to the actual proportions of cell types in the source tissues, but the relationship is imprecise. More reliable quantitative information about neuron-type proportions is likely to emerge from ongoing spatial transcriptomics studies and is certain to be useful but unlikely to be critical to the present analysis. Panels (A) and (B) also introduces the Tasic neuron-type-coding color scheme for glutamatergic neurons, which are used consistently in Figs. 3-7 of the main text. The table in (B) enumerates and labels each of the 55 glutamatergic neuron type leaf clusters graphed in (A). These neuron type labels were generated by Tasic, et al., to capture salient distinguishing feature for each neuron type. For glutamatergic types, these features always include cortical layer and a VISp vs ALM area designation wherever a cell type was found only in that one area. The labels also include names of key marker genes. Panel (B) shows that most of the 55 Tasic glutamatergic types are exclusive either to VISp or to ALM, with only four (*41. L6 CT Nxph2 Wls, 48. L6b Col8a1 Rxfp1, 53. L6b P2ry12 and 55. L6b Hsd17b*) common to both areas.

Panels (C) and (D) represent the Tasic 2018 taxonomy of 60 GABAergic cortical neuron types just as (A) and (B) represented the 55 Glutamatergic types. These figures also introduce the Tasic type-coding color scheme for GABAergic neurons, which again will be used consistent in all following figures. As evident from comparison of (B) and (D), and discussed extensively by Tasic, et al., GABAergic neuron types are much more prevalently common to both VISp and ALM cortical areas, with all but two (*88. Sst Tac1 Tacr3 and 112. Pvalb Reln Itm2a*) of the 60 GABAergic types found in both areas. For GABAergic neurons, the Tasic type labels refer strictly to marker gene combinations. A notably large fraction of the marker genes appearing in the Tasic type labels are either NPP or NP-GPCR genes. Names of these genes are highlighted by bold red type in both (B) and (D). Most of these red-highlighted NP genes can be found in Table 1 but a few cannot because they did not meet our curation criterion of high cortical expression of both members of a conjugate NPP and NP-GPCR pair.

Panel (E) concatenates the Tasic glutamatergic and GABAergic neuron-type rosters and type-code color strips reproduced in Figs. 3A-D, tailored as needed to represent (i) all 115 neurons types in VISp and ALM, (ii) the 93 VISp-only types and (iii) the 84 ALM-only types. These edited rosters and code strips maintain a neuron type sequence dictated by the cladograms of (A) and (C). Though the sequence of glutamate and GABA types is somewhat arbitrary, we concatenated the rosters using glutamatergic-first ordering used by Tasic, et al. [32]. Note that (E) sub-panels i, ii and iii are aligned laterally to keep the conserved GABAergic cell types in horizontal register. Type-color pixels are replaced with black spacers for the two GABAergic cell types that are not present in ALM, again to maintain horizontal register of the conserved GABAergic types. Because few glutamatergic cell types are conserved between VISp and ALM, there is no such precise register for the glutamatergic types. The annotated color strips in this panel are provided to aid in interpretation of subsequent figures where space and font-size constraints render repeated typographic annotation impractical. It is hoped that even individuals with limited color vision will find these type-code color strips useful due to the redundancy of type order, luminance and hue information.

Supplementary Methods

Autoencoder classification development and evaluation methods. This work used two types of gene datasets: 1) The "HE" gene set, which contains the expression of 6083 highly expressed neuronal genes in 22,439 neurons and 2) 47 gene sets, which contain the expression of 47 specific genes in 22,439 neurons. Both "HE" and 47 gene datasets are divided into training and validation sets using a 92%-8% splits.

Autoencoders are a deep neural network model that consist of two subnetworks: an encoder subnetwork to compress high dimensional inputs into a low dimensional representation, and a decoder subnetwork to estimate the input from the low dimensional representation. We constructed a network with two autoencoders, with 8 hidden layers each. The first autoencoder, referred as "HE gene autoencoder", has an input layer with 6083 units, with 20% of the inputs randomly and independently removed using a dropout operation. The second autoencoder has an input layer with 47 units. (the smaller architecture was chosen to reflect the size of the NP gene set.) In both autoencoders, the hidden units use the rectified linear function as the non-linear transformation, while the latent units use batch normalization. The networks were trained using the mean squared error between the input and the output layers as the loss function, using the Adam optimizer and a batch size of 956. Implementations were in Python using Tensorflow [119] and Keras [120] libraries on an NVIDIA TitanX.

We started with training the HE gene autoencoder for 50,000 epochs using the HE gene set to minimize the reconstruction error for the gene expression levels. To determine the latent space dimensionality for optimal HE gene representation, we performed hyperparameter tuning by varying the latent space dimensionality between 2 and 20 dimensions. The performance of each latent space was compared to the Tasic 2018 taxonomy using a supervised gaussian mixture model (GMM). An optimal dimensionality of 5 was chosen, which resulted in a peaking adjusted random index of 0.8672. After training the HE autoencoder, the second autoencoder with the smaller architecture was trained using one of the sets of 47 genes for 10,000 epochs. The smaller autoencoder was trained to minimize the reconstruction error, while attempting to match the representation learned based on the HE gene set. The same procedure was used for all small gene subsets including NP and random gene sets.

To quantify the similarity between cell-type clustering from the autoencoder latent space representation and the Tasic 2018 taxonomy, progressively simplified taxonomies were trained by iterative merging, and the most complex taxonomy for which each cell which was correctly mapped was recorded as the Resolution Index (RI). RI measures the depth of the lowest common ancestor of the predicted node and original node in the taxonomy, from the lowest resolution at the root (RI = 0) to the finest resolution at the leaves (RI = 1). Cell-type class labels for clustering were initialized with leaf nodes from the Tasic 2018 hierarchical tree. The number of class labels was reduced by 1 in each iteration, by merging the lowest pair of nodes on the hierarchical tree. Cells for which the class label was correctly assigned by supervised clustering were assigned the RI associated with the corresponding node, and the procedure was continued until all cells were assigned an RI score. (All cells were used as input in the GMM thus predictions were also assigned to all cells.) To account for the exclusion of all non-neuronal cell-types, scores were normalized over the resolution index corresponding to the first neuronal node on the Tasic 2018 taxonomy [32]. The performance of each gene set was quantified by taking the average RI score for all cells from the respective latent space representation.