Single-Cell Transcriptomic Evidence for Dense Intracortical Neuropeptide Networks

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

Allen Institute for Brain Science, 615 Westlake Ave N, Seattle, WA 98109, USA
* Neuroscience Paris Seine, Sorbonne Université, 9 Quai Saint Bernard, 75005 Paris, France

Impact

Single-cell mRNA sequencing data from mouse neocortex expose evidence for peptidergic neuromodulatory networks that locally interconnect every cortical neuron

Data Highlights

- At least 97% of mouse neocortical neurons express one or more of 18 neuropeptide precursor proteins (NPP) genes at very high levels
- At least 98% of cortical neurons express one or more of 29 neuropeptide-selective G-protein-coupled receptor (NP-GPCR) genes cognate to the 18 highly expressed NPP genes
- Neocortical expression of these 18 NPP and 29 NP-GPCR genes is highly neuron-type-specific and their expression patterns differentiate transcriptomic neuron types with exceptional power
- Neuron-type taxonomy and type-specific expression of 37 cognate NPP / NP-GPCR gene pairs generate testable predictions of at least 37 local intracortical neuromodulation networks

Summary

Seeking new insights into the homeostasis, modulation and plasticity of cortical synaptic networks, we have analyzed results from a single-cell RNA-seq study of 22,439 mouse neocortical neurons. Our analysis exposes transcriptomic evidence for dozens of molecularly distinct neuropeptidergic modulatory networks that directly interconnect all cortical neurons. This evidence begins with a discovery that transcripts of one or more neuropeptide precursor (NPP) and one or more neuropeptide-selective G-protein-coupled receptor (NP-GPCR) genes are highly abundant in all, or very nearly all, cortical neurons. Individual neurons express diverse subsets of NP signaling genes from palettes encoding 18 NPPs and 29 NP-GPCRs. These 47 genes comprise 37 cognate NPP/NP-GPCR pairs, implying the likelihood of local neuropeptide signaling. Here we use neuron-type-specific patterns of NP gene expression to offer specific, testable predictions regarding 37 peptidergic neuromodulatory networks that may play prominent roles in cortical homeostasis and plasticity.

Correspondence:

Stephen J Smith, stephens@alleninstitute.org; Michael Hawrylycz, mikeh@alleninstitute.org

13,763 words (Introduction, Results, Discussion, Methods, in-line citations and figure legends)

8 Figures; 16 Figure Supplements; 134 references

Keywords: transcriptomes, neuropeptides, mouse, neocortex, visual cortex, motor cortex, neuron types, neuron-type taxonomy, neurotaxonomy, mRNA-seq, GPCR, synaptic plasticity, neuromodulation, homeostasis, synaptic network homeostasis

Introduction

Neuromodulation - the graded and relatively slow adjustment of fast synapse and ion channel function via diffusible cell-cell signaling molecules - is a fundamental requirement for adaptive nervous system function (Abbott and Regehr, 2004; Bargmann, 2012; Bucher and Marder, 2013; Katz and Lillvis, 2014; Marder, 2012; Marder et al., 2015; Mccormick and Nusbaum, 2014; Nadim and Bucher, 2014; Nusbaum et al., 2017). Neuromodulator molecules take many different chemical forms, including diatomic gases such as nitric oxide, lipid metabolites such as the endocannabinoids, and amino acids and their metabolites such as glutamate, GABA, acetylcholine, serotonin and dopamine. By far the largest family of neuromodulator molecules known, however, comprises the evolutionarily ancient proteinaceous signaling molecules known as neuropeptides (Baraban and Tallent, 2004; Burbach, 2011; Gonzalez-Suarez and Nitabach, 2018; Hökfelt et al., 2013; van den Pol, 2012; Wang et al., 2015). The most widely studied neuropeptides are the endogenous “opioid” peptides - enkephalins, endorphins and dynorphins - but there are nearly one hundred other NPP genes in the human genome and numerous homologs are present in almost all known animal genomes (Elphick et al., 2018; Jekely, 2013).

The broadest definition of “neuropeptide” would embrace any soluble peptide that serves as a messenger by diffusing from one neuron to another. A narrower but more common definition (Burbach, 2011) requires (1) translation of a neuropeptide precursor protein (NPP) into the lumen of a source neuron’s rough endoplasmic reticulum (rER), (2) enzymatic cleavage of the NPP into one or more neuropeptide (NP) products during or after passage through the rER–Golgi complex and packaging into dense-core vesicles, (3) transport and storage of dense-core vesicles within the source neuron, (4) secretion of the NP product upon demand by activity- and calcium-dependent dense-core vesicle exocytosis, and then (5) interstitial diffusion to act upon a target neuron by binding to a ligand-specific receptor. This pathway enlarges the potential palette of distinct neuropeptides beyond that established simply by the large number of NPP genes, as a given NPP may be differentially cleaved into alternative NP products during its intracellular and interstitial passage.

Most neuropeptide receptors are encoded by members of the very large superfamily of G-protein-coupled receptor (GPCR) genes (Hoyer and Bartfai, 2012; Krishnan and Schioth, 2015; Mains and Eipper, 2006; van den Pol, 2012). GPCRs are selective, high-affinity receptors distinguished by characteristic seven-transmembrane-segment atomic structures and signal transduction involving heterotrimeric G-proteins (hence their name). Genes encoding GPCRs selective for neuropeptides (NP-GPCR genes) and for most of the other chemical neuromodulators mentioned above are found in the genomes of almost all metazoans: phylogenomic evidence suggests very early evolutionary origins (Jekely, 2013; Katz and Lillvis, 2014), possibly even predating evolution of the synapse (Varoqueaux and Fasshauer, 2017). In modern metazoan nervous systems, synapses rely primarily upon recycling small molecule transmitters and ligand-gated ion channels (alternatively known as “ionotropic receptors”) for fast (millisecond timescale) transmission, but GPCRs selective for widely varied ligands including the fast recycling transmitters and many other secreted molecules (e.g., glutamate, GABA, acetylcholine, monoamines and neuropeptides) play critical roles in the slower modulation of fast synaptic transmission and electrical activity (Elphick et al., 2018; Grimmelikhuijzen and Hauser, 2012; Jekely, 2013; Krishnan and Schioth, 2015; Varoqueaux and Fasshauer, 2017).

Because modulatory neuropeptides are not subject to the rapid transmitter re-uptake and/or degradation processes necessary for fast synaptic transmission, secreted neuropeptides are thought to persist long enough (e.g., minutes) in brain interstitial spaces for diffusion to very-high-affinity NP-GPCRs hundreds of micrometers distant from release sites (Ludwig and Leng, 2006; Nässel, 2009; Russo, 2017). Neuropeptide signaling can thus be presumed both “paracrine”, with secretion from individual neurons hitting receptor-positive cells over substantial diffusion distances and converging by diffusion from many local secreting neurons onto single receptor-positive neurons, and to be relatively slow (seconds-to-minutes timescale of action). Though present information is limited, eventual degradation by interstitial

peptidases nonetheless probably restricts diffusion of most neuropeptides to sub-millimeter, local circuit distance scales.

The receptors encoded by NP-GPCR genes are highly diverse in ligand specificity but less diverse in downstream signaling impacts. Although GPCR signaling has long been recognized as complex and many faceted (Hamm, 1998), most known neuronal NP-GPCR actions reflect phosphorylation of ion channels, synaptic proteins or transcription factors mediated by protein kinases dependent on the second messengers cyclic AMP or calcium (Mains and Eipper, 2006; Nadim and Bucher, 2014; van den Pol, 2012). Primary effects of NP-GPCRs expressed in cortex, in turn, fall into just three major categories distinguished by G-protein alpha subunit (G_α) family. The $G_{i/o}$ family (i/o) inhibits cAMP production, the G_s family (s) stimulates cAMP production, and the $G_{q/11}$ family (q) amplifies calcium signaling dynamics (Syrovatkina et al., 2016). For most NP-GPCR genes, the primary G_α family (e.g., i/o, s or q/11) is now known (Alexander et al., 2017) and offers a good first-order prediction of the encoded GPCR's signal transducing action. The profound functional consequences of neuromodulation by GPCRs range from modification of neuronal firing properties and calcium signaling dynamics through regulation of synaptic weights and synaptic plasticity (Bargmann, 2012; Markram et al., 2013; McCormick and Nusbaum, 2014).

It is well established that certain neuropeptides, including vasoactive intestinal peptide (VIP), somatostatin (SST), neuropeptide Y (NPY), substance P, and cholecystokinin (CCK), are detectible at high levels in particular subsets of GABAergic cortical neurons (Tremblay et al., 2016). These neuropeptides have come into broad use as markers for particular GABAergic interneuron classes, while the corresponding NPP and NP-GPCR genetics have provided molecular access to these and other broad neuron type classes (Daigle et al., 2018; Maximiliano José et al., 2018). *In situ* hybridization and microarray date, e.g., the Allen Brain Atlases (Hawrylycz et al., 2012; Lein et al., 2007), have also established that mRNA transcripts encoding these five NPPs and that many other NPPs and NP-GPCR genes are expressed differentially in many brain regions. There has been a critical lack, however, of comprehensive expression data combining whole-genome depth with single-cell resolution. Absent such data, it has been difficult to generate specific and testable hypotheses regarding cortical neuropeptide function and to design robust experiments to test those hypotheses (Tremblay et al., 2016; van den Pol, 2012).

Here we describe new findings regarding NPP and NP-GPCR gene expression in mouse cortex. These findings have surfaced during a focused analysis of a previously published single-cell RNA-seq data acquired from large numbers of isolated mouse cortical neurons (Tasic et al., 2018). We begin by leveraging only the genomic depth and single-cell resolution of this dataset. Next, we briefly introduce the transcriptomic neurotaxonomy developed in the same resource publication and explore the additional analytical power of a neurotaxonomic framework. Then, we distill these findings into specific and testable predictions concerning intracortical peptidergic modulation networks. Finally, we discuss the potential of a neurotaxonomically integrated view of neuromodulatory and synaptic networks to reveal previously obscure principles of cortical sensory, mnemonic and motor function.

Results

The present study is based on analysis of a resource single-cell RNA-seq dataset acquired at the Allen Institute (Tasic et al., 2018) and available for download at *http://celltypes.brain-map.org/rnaseq/.* These RNA-seq data were acquired from a total of 22,439 isolated neurons, with detection of transcripts from a median of 9,462 genes per cell (min = 1,445; max = 15,338) and an overall total of 21,931 protein-coding genes detected. 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). Single neuron harvesting methods were designed to mildly enrich samples for GABAergic neurons, such that the sampled neuron population is roughly half GABAergic (47%) and half glutamatergic (53%). The resource publication (Tasic et al., 2018) should be consulted for full details of neuron harvesting, sample preparation, sequencing and data processing. Since we refer very frequently here to this resource publication and dataset, we'll refer to both now simply as "Tasic 2018", and all further references here to

neuron "class", "subclass" or "type" should be understood to refer specifically to the particular mouse neocortex neurotaxonomy described in the Tasic 2018 publication.

The Tasic 2018 single-cell RNA-seq data tables report the abundance of transcripts from individual neurons in both "counts per million reads" (CPM) and "fragments per kilobase of exon per million reads mapped" (FPKM) units. Our analysis of this data compares gene expression levels quantitatively, with two distinct use cases: (1) comparisons across large sets of different genes, and (2) comparisons of the same gene across different individual cells, cell types and brain areas. We have relied upon FPKM data (Mortazavi et al., 2008; Pimentel, 2014), for use case 1 (i.e., the Table 1 and 2 comparisons across genes). For use case 2 (as in all figures below), we have preferred the CPM units, because these units were used to generate the Tasic 2018 neurotaxonomy. In any case, choice of CPM vs. FPKM units would have very little impact on the present outcomes.

NPP Gene	Peak FPKM	pFPKM Percentile	pFPKM Rank	% Cells	Predicted Neuropeptides	Cognate NP-GPCR Genes
Npy	108,865	100.00	1	42	Neuropeptide Y	*Npy1r, Npy2r, Npy5*
Sst	70,274	99.99	2	26	Somatostatins	*Sstr1, Sstr2, Sstr3, Sstr4*
Vip	48,747	99.99	3	33	Vasoactive Intestinal Peptide	*Vipr1, Vipr2*
Tac2	18,284	99.98	4	15	Neurokinin B	*Tacr3*
Cck	16,396	99.97	6	69	Cholecystokinins	*Cckbr*
Penk	11,160	99.96	8	26	Enkephalins	*Oprd1, Oprm1*
Crh	9,118	99.95	10	17	Corticotropin-Releasing Hormone	*Crhr1, Crhr2*
Cort	7,477	99.93	15	32	Cortistatin	*Sstr1, Sstr2, Sstr3, Sstr4*
Tac1	5,728	99.92	18	11	Substance P, Neurokinin A	*Tacr1*
Pdyn	2,813	99.69	68	8	Dynorphins	*Oprd1, Oprk1, Oprm1*
Pthlh	1,656	99.29	156	18	Parathyroid-Hormone-Like Hormone	*Pth1r*
Pnoc	698	97.68	509	23	Nociceptins	*Oprl1*
Trh	510	96.51	766	3	Thyrotropin-Releasing Hormone	*Trhr, Trhr2*
Grp	435	95.59	968	12	Gastrin-Releasing Peptide	*Grpr*
Rln1	258	91.99	1757	7	Relaxin 1	*Rxfp1, Rxfp2, Rxfp3*
Adcyap1	165	87.29	2788	26	Adenylate Cyclase-Activating Polypeptides	*Adcyap1r1, Vipr1, Vipr2*
Nts	121	82.14	3917	1	Neurotensin	*Ntsr1, Ntsr2*
Nmb	112	80.53	4270	14	Neuromedin B	*Nmbr*

Table 1. Single-cell RNA-seq expression statitics for 18 highly expressed neuropeptide precursor protein (NPP) genes cognate to locally expressed NP-GPCR genes (see Table 2). NPP genes are tabulated here along with peak single-cell expression levels as pFPKM (Peak FPKM, see **Materials and Methods**), percentile and absolute ranking of these pFPKM values across pFPKMs for all 21,931 protein-coding genes, and the percentage of cells sampled in which transcripts of the given NPP gene were detected at > 1 CPM. The table also lists predicted neuropeptide products, and genes encoding the locally expressed G-protein-coupled receptors (NP-GPCRs) cognate to the NPP (see Table 2). NPP genes are listed here in descending order of Peak FPKM. Pastel color fills in the "Cognate NP-GPCR Genes" column correspond to i/o (pink), s (light green) and q/11 (light blue) transduction families of associated G-protein and will be used to highlight these families consistently in all following figures.

Single-neuron expression profiles of 18 select neuropeptide precursor (NPP) genes

Table 1 lists results of analyzing the expression of 18 NPP genes in all 22,439 individual neurons represented in Tasic 2018. Here we have made use of the "peak FPKM" (pFPKM) metric described in **Materials and Methods** below to quantify the expression of specific genes in highly expressing subsets of single-cell populations that exhibit highly variant expression of that particular gene. Each of the 18 NPP genes on this list meets two conditions: (1) the included NPP gene is highly expressed (top quintile pFPKM over all protein-coding genes) across VISp and ALM cortical areas, and (2) at least one gene for an NP-GPCR selective for the predicted product of at least one of the 18 NPPs is highly expressed in neurons within the same local area of neocortex (see Table 2). The first requirement was imposed to increase the likelihood of active secretion of the NP product encoded by the candidate NPP gene. The second requirement, for "cognate" pairing between each included NPP and a locally expressed NP-GPCR gene,

was imposed to elevate the likelihood of paracrine NP signaling within a cortical local circuit volume, as envisioned in **Introduction** above. The process for selection of these 18 NPP genes is described in more detail in **Materials and Methods**. **Table 1** lists Peak FPKM values for each of the 18 NPP genes, percentile and absolute ranks of that Peak FPKM value across all protein-coding genes, the percentage of cells sampled in which expression of the listed gene is detectible, predicted neuropeptide product(s) encoded, and the NP-GPCR gene(s) fulfilling requirement (2) for that NPP gene. Gene ontology results for the 18 select NPP genes are provided by **Supplementary file 1**.

The Peak FPKM ranking columns in **Table 1** show that expression levels of most of the 18 NPP genes are extremely high in the range of values for all 21,931 protein-coding genes detected in all 22,439 neurons sampled. Of these genes, *Npy, Sst, Vip* and *Tac2* rank as the top four overall in pFPKM values, while *Cck, Penk* and *Crh* also rank in the top ten. Eleven of these NPP genes rank in the top percentile and all 18 rank in the top quintile by pFPKM. To the simplest first approximation, very high abundance of a given protein-coding transcript implies the potential, at least, for a very high rate of synthesis of the encoded protein. The extremely high peak abundance of these NPP transcripts thus suggests that NP precursor proteins could be synthesized at very high rates in neurons exhibiting such peak abundance. In a steady state, a high rate of synthesis wouild then necessarily imply a correspondingly high overall rate of protein product elimination. For an NP precurson protein, processing and secretion of active neuropeptide would seem the obvious and most likely route of elimination. The high abundance of transcripts encoding these 18 NPPs might thus be construed as *prima facie* evidence for robust secretion of neuropeptide products.

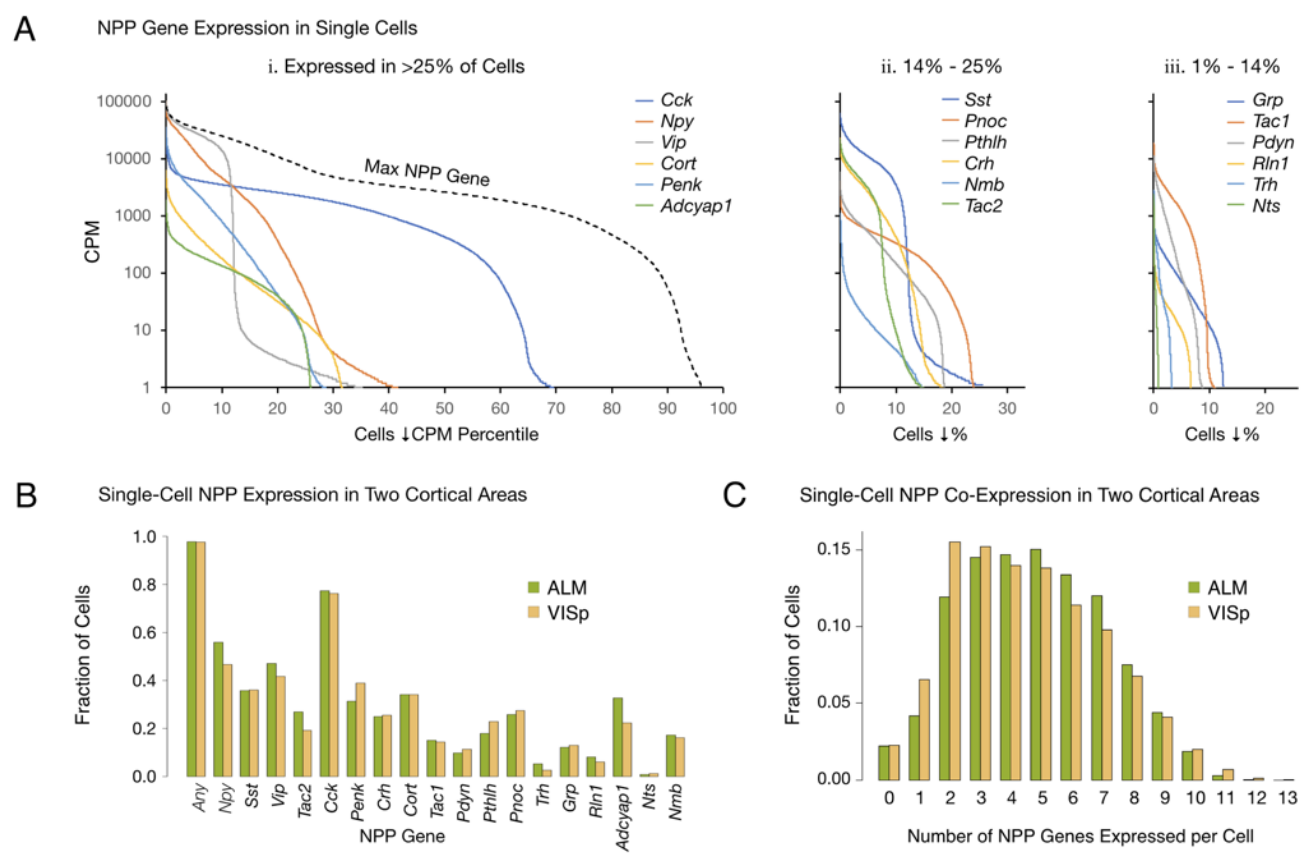


Figure 1. Single-cell NPP gene expression and co-expression statistics for distant neocortical areas VISp and ALM show that expression patterns for 18 NPP genes are highly differential within both neocortical areas but highly conserved between areas. (A) Different NPP genes show very different expression level distributions across the 22,439 VISp+ALM neurons sampled. Color-coded solid curves plot single-cell CPM values for the specified individual NPP genes in descending order along a cell population percentile axis. The 18 curves are segregated for clarity into three panels (I, ii, iii) sorted by cell population

percentiles at which CPM values fall below 1. Large differences in fractions of cells expressing different NPP genes are evident. The dashed curve labeled "Max NPP Gene" in panel A.i was generated by plotting CPM values of for the most abundant NPP transcript in 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 detailed in panel A. (C) Histograms illustrating frequencies of various multiples of NPP genes co-expressed 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.

The following source data are available for figure 1:

Source data 1. Sorted NPP CPM distributions for all neurons.

Source data 2. Single-cell NPP CPM expression table.

Figure 1A quantifies differential expression of the 18 NPP genes listed in **Table 1**. Each of the 18 color-coded solid curves represents a distribution of single-neuron CPM values for one NPP gene. Curves were generated by plotting CPM for each individual neuron in descending rank order along a cell population percentile axis. Each curve exhibits a transition from high to very low (commonly zero) expression across the sampled neuron population, but these transitions occur over very different population percentile ranges, providing clear evidence for highly differential single-cell expression of these genes. Percentages of the sampled neuron population with detectible expression of a given NPP gene range from more than 65% for *Cck* down to ~1% for *Nts*. (Note, however, that the cell population sampled has been enriched for GABAergic cell types as described in Tasic 2018).

The RNA-seq data suggest that all, or at least very nearly all, neocortical neurons express at least one NPP gene. The dashed curve in **Figure 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. When one takes into account the pulsatile nature of transcription (Suter et al., 2011) and the stochastic nature of RNA-seq transcript sampling (Fu and Pachter, 2016; Kim et al., 2015; Tasic et al., 2016), these numbers might be best understood as lower limits. The results summarized in **Figure 1A** may therefore be consistent with the proposition that every cortical neuron is peptidergic.

Conservation of NPP gene expression statistics between VISp and ALM

The paired bars in **Figure 1B** represent fractions of cells expressing a given gene in each of the two cortical areas. It is obvious that the differential expression profiles in VISp and ALM are highly similar (ρ=0.972, p<1.72E-11), 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 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. **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. **Figure 1C** also reveals strong similarities of NPP co-expression distributions between VISp and ALM.

Single-neuron expression profiles of 29 select neuropeptide receptor (NP-GPCR) genes

Table 2 lists 29 NP-GPCR genes that are highly expressed in varied subsets of the 22,439 individual neurons sampled from cortical areas VISp and ALM. These 29 genes encode receptor proteins substantially selective for neuropeptide products encoded by the 18 NPP genes of **Table 1** (cross-referenced from 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 A-F GPCR class and expected primary G_α family and cross-references back to 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 pairing and the consequent implication of local intracortical signaling. Methods of NP-GPCR gene selection are described more fully in **Materials and Methods**. A more complete listing of NP-GPCR genes with pFPKM values in provided by Table 2 source data 1. Gene ontology results for the 29 select NPP genes are provided by **Supplementary file 2**.

The "pFPKM Percentile" column in **Table 2** shows that most of these 29 NP-GPCR genes are expressed in cortex with Peak FPKM values well above median (50th percentile) for all protein coding genes. The range of cortical neuron pFPKM values for NP-GPCR genes does not match the extreme heights noted for NPP genes, but this is as expected given that NP-GPCR gene products are thought to be durable cellular components, unlikely to be rapidly disposed by secretion as expected for NPP gene products. Peak FPKM values for NP-GPCR transcripts are nonetheless quite high in the range of transcripts of other durable cellular component genes, suggesting a strong likelihood that they are indeed translated into functionally important protein products.

NP-GPCR Gene	Peak FPKM	pFPKM Percentile	% Cells	Neuropeptide Receptor	GPCR Class	Primary G_α Family	Cognate NPP Genes
Sstr2	413	95.3	42	Somatostatin Receptor 2	A4	$G_{i/o}$	*Sst, Cort*
Npy2r	291	93.1	10	Neuropeptide Y Receptor Y2	A9	$G_{i/o}$	*Npy*
Npy1r	272	92.4	50	Neuropeptide Y Receptor Y1	A9	$G_{i/o}$	*Npy*
Grpr	231	91	10	GRP Receptor	A7	$G_{q/11}$	*Grp*
Cckbr	210	90	52	Cholecystokinin B Receptor	A6	$G_{q/11}$	*Cck*
Ntsr2	161	86.9	17	Neurotensin Receptor 2	A7	$G_{q/11}$	*Nts*
Npy5r	152	86.1	28	Neurpeptide Y Receptor Y5	A9	$G_{i/o}$	*Npy*
Nmbr	123	82.4	8	Neuromedin B Receptor	A7	$G_{q/11}$	*Nmb*
Rxfp1	121	82	22	Relaxin Family Receptor 1	A5	G_s	*Rln1*
Sstr4	106	79.5	28	Somatostatin Receptor 4	A4	$G_{i/o}$	*Sst, Cort*
Trhr	101	78.4	3	TRH Receptor	A7	$G_{q/11}$	*Trh*
Sstr1	90	76	38	Somatostatin Receptor 1	A4	$G_{i/o}$	*Sst, Cort*
Adcyap1r1	89	75.8	71	ADCYAP1 Receptor 1	B1	G_s	*Adcyap1, Vip*
Crhr1	86	74.9	28	CRH Receptor 1	B1	G_s	*Crh*
Rxfp3	85	74.7	5	Relaxin Family Receptor 3	A5	$G_{i/o}$	*Rln1*
Oprl1	82	73.8	48	Opioid Receptor-Like 1	A4	$G_{i/o}$	*Pnoc*
Crhr2	72	70.7	3	CRH Receptor 2	B1	G_s	*Crh*
Tacr3	65	68	3	Tachykinin Receptor 3	A9	$G_{q/11}$	*Tac2*
Oprk1	64	67.4	3	Kappa-Opioid Receptor	A4	$G_{i/o}$	*Pdyn*
Tacr1	56	64.2	3	Tachykinin Receptor 1	A9	$G_{q/11}$	*Tac1*
Pth1r	51	61.6	15	PTH 1 Receptor	B1	$G_{q/11}$	*Pthlh*
Vipr1	41	56.1	28	VIP Receptor 1	B1	G_s	*Vip, Adcyap1*
Oprm1	35	52.1	43	Mu-Opioid Receptor	A4	$G_{i/o}$	*Penk, Pdyn*
Trhr2	30	48.9	10	TRH Receptor 2	A7	$G_{q/11}$	*Trh*
Vipr2	30	48.4	0.5	VIP Receptor 2	B1	G_s	*Vip, Adcyap1*
Rxfp2	28	47.3	4	Relaxin Family Receptor 2	A5	G_s	*Rln1*
Oprd1	26	45.8	13	Delta-Opioid Receptor	A4	$G_{i/o}$	*Penk, Pdyn*
Ntsr1	24	44.3	10	Neurotensin Receptor 1	A7	$G_{q/11}$	*Nts*
Sstr3	17	39.5	21	Somatostatin Receptor 3	A4	$G_{i/o}$	*Sst, Cort*

Table 2. Single-cell RNA-seq expression statitics for 29 neuropeptide-selective, G-protein-coupled receptor (NP-GPCR) genes cognate to locally expressed NPP genes (see Table 1). NP-GPCR gene peak FPPM values, percentile ranking, and percentage sampled as for NPP genes in Table 1. The table names encoded NP-GPCR proteins, A-F class of NP-GPCR, primary G_α signal transduction family (Alexander et al., 2017) and cognate NPP genes. Pastel color fill in "primary G_α family" column as in Table 1.

The single-cell RNA-seq data expose very highly differential expression of NP-GPCR genes in cortical neurons. Figure 2 represents expression patterns of the 29 NP-GPCR genes listed in Table 2 in the same manner as for the 18 NPP genes in Figure1. Figure 2A establishes that each of the 29 NP-GPCR genes is expressed in highly differential fashion across the 22,439 mouse cortical neurons sampled, as was the case for the 18 NPP genes. As was noted for NPP genes in **Figure 1**, each of the curves in Figure 2A exhibits a transition from very high to very low (commonly zero) expression across the sampled neuron population. These transitions occur at very different population percentile points, again providing clear evidence for highly differential expression. 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*.

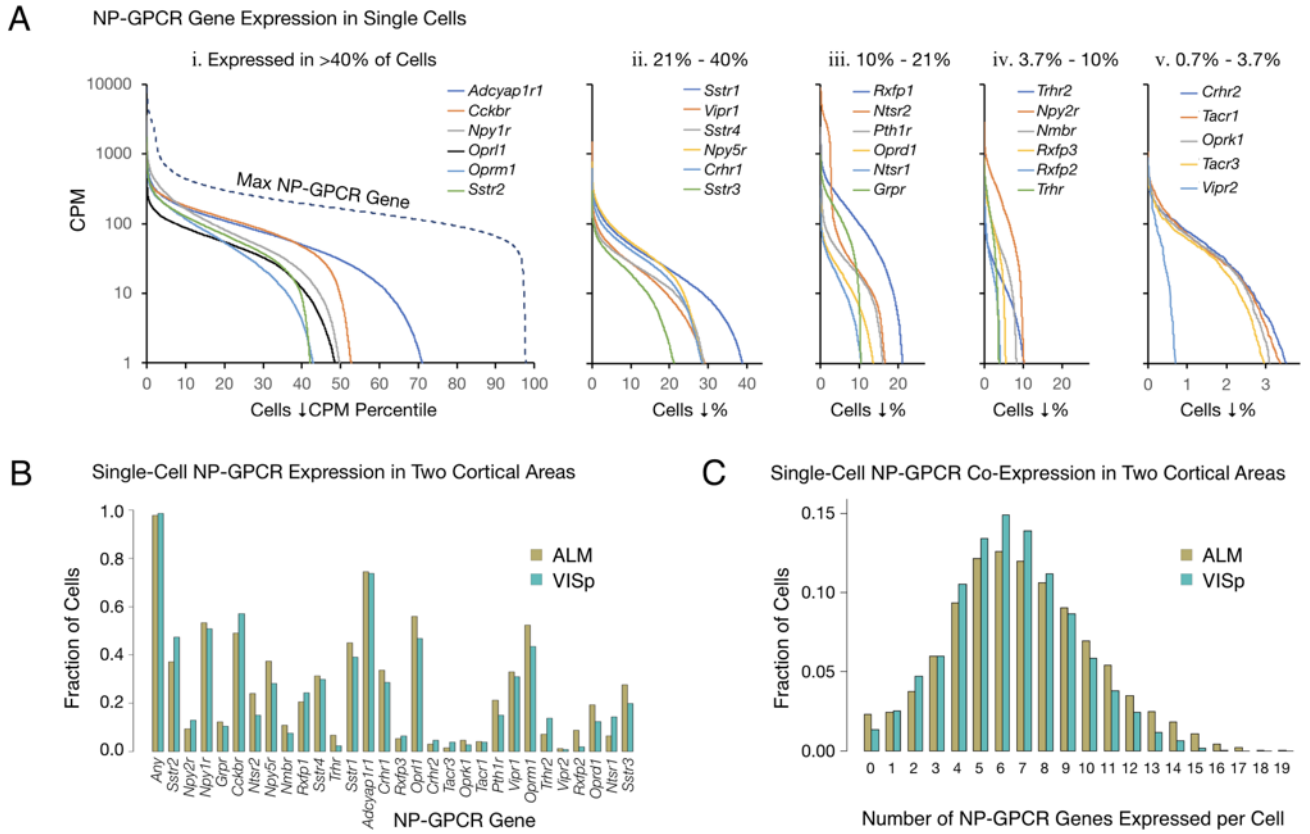


Figure 2. Single-cell NP-GPCR gene expression and co-expression statistics for distant neocortical areas VISp and ALM show that expression patterns for 29 NP-GPCR genes are highly differential within neocortical areas but conserved between areas. (A) Different NP-GPCR genes show very different expression level distributions across the 22,439 VISp+ALM neurons sampled. Color-coded solid curves plot single-cell CPM values for the specified individual NP-GPCR genes in descending order along a cell population percentile axis. The 29 curves are segregated for clarity into five panels (i-v) 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 labeled “max NP-GPCR Gene” in panel A.i 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.

The following figure supplements and source data are available for figure 2:

Figure supplement 1. Co-expression of cognate NPP/NP-GPCR pairs.

Source data 1. Sorted NP-GPCR CPM distributions for all neurons.

Source data 2. Single-cell NP-GPCR CPM expression table.

The RNA-seq data suggest that all, or at least very nearly all, neocortical neurons express at least one NP-GPCR gene. The dashed curve in the left panel of Figure 2A, generated similarly to the dashed curve for NPP genes in **Figure 1**A, 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 NPP curve in **Figure 1**, but still indicative of quite high expression. 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 Figure 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.

Conservation of NP-GPCR gene expression statistics between VISp and ALM

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, again revealing strong similarities of differential expression profiles in the two very different neocortical areas (ρ=0.959, p<2.2E-16). 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. As illustrated by Figure 2 – Figure supplement 1, it is furthermore common for individual neurons to co-express cognate NPP/NP-GPCR pairs, raising the intriguing possibility of cell-autonomous feedback mediated by an autocrine action of a secreted NP product on the secreting cell itself. Figure 2 – Figure supplement 1 additionally shows that cognate pair co-expression patterning is also highly conserved between areas VISp and ALM.

Neurotaxonomic profiling of NPP and NP-GPCR gene expression

The analysis so far has relied solely upon the genomic depth and single-cell resolution characteristics of the Tasic 2018 transcriptomic data. We now proceed to explore the analytical power of the transcriptomic neurotaxonomy developed as part of the Tasic 2018 study. This neurotaxonomy makes it possible to predict a protein "parts list" for any neuron that can be mapped to a given transcriptomic taxon. Combined with tools for genetic access to transcriptomic taxa, transcriptomic taxonomy thereby offer rich prospects for experimental test of such predictions (see also **Discussion** below), The present analysis will make extensive use of the Tasic 2018 neurotaxonomy's representation of 115 glutamatergic and GABAergic transcriptomic neuron types (see **Figure 3 - figure supplement 1**).

Figure 3 shows transcriptomic gene expression "heatmaps", representing transcript abundance for each of 18 NPP (**Figure 3A**) and 29 NP-GPCR (**Figure 3B**) across all 115 glutamatergic and GABAergic neuron types by $\log_{10}$-scaled pseudocolor. These heatmaps show that expression of every one of these 47 genes is highly specific to particular neuron types, but that type specificity varies greatly from gene to gene. Note that CPM expression levels vary across neuron types by factors exceeding 10,000 for many NPP genes and 1,000 for many NP-GPCR genes. These heat maps also show every neuron type expresses multiple NPP and NP-GPCR genes and that each of the NPP and NP-GPCR genes is expressed in multiple neuron types (with Vipr2 in one Pvalb type as a near exception). These two heatmaps further show many cases where both an NPP gene and its cognate NP-GPCR receptor are expressed in the same neuron type, with the *Cck* / *Cckbr* and *Adcyap1* / *Adcyap1r1* pairs being particularly prominent examples. Quite intriguingly, these expression heat maps also suggest that each of the neuron types

might be distinguished by a unique pattern of expression of these 47 NP genes. This possibility will be explored quantitatively in connection with Figure 4 below.

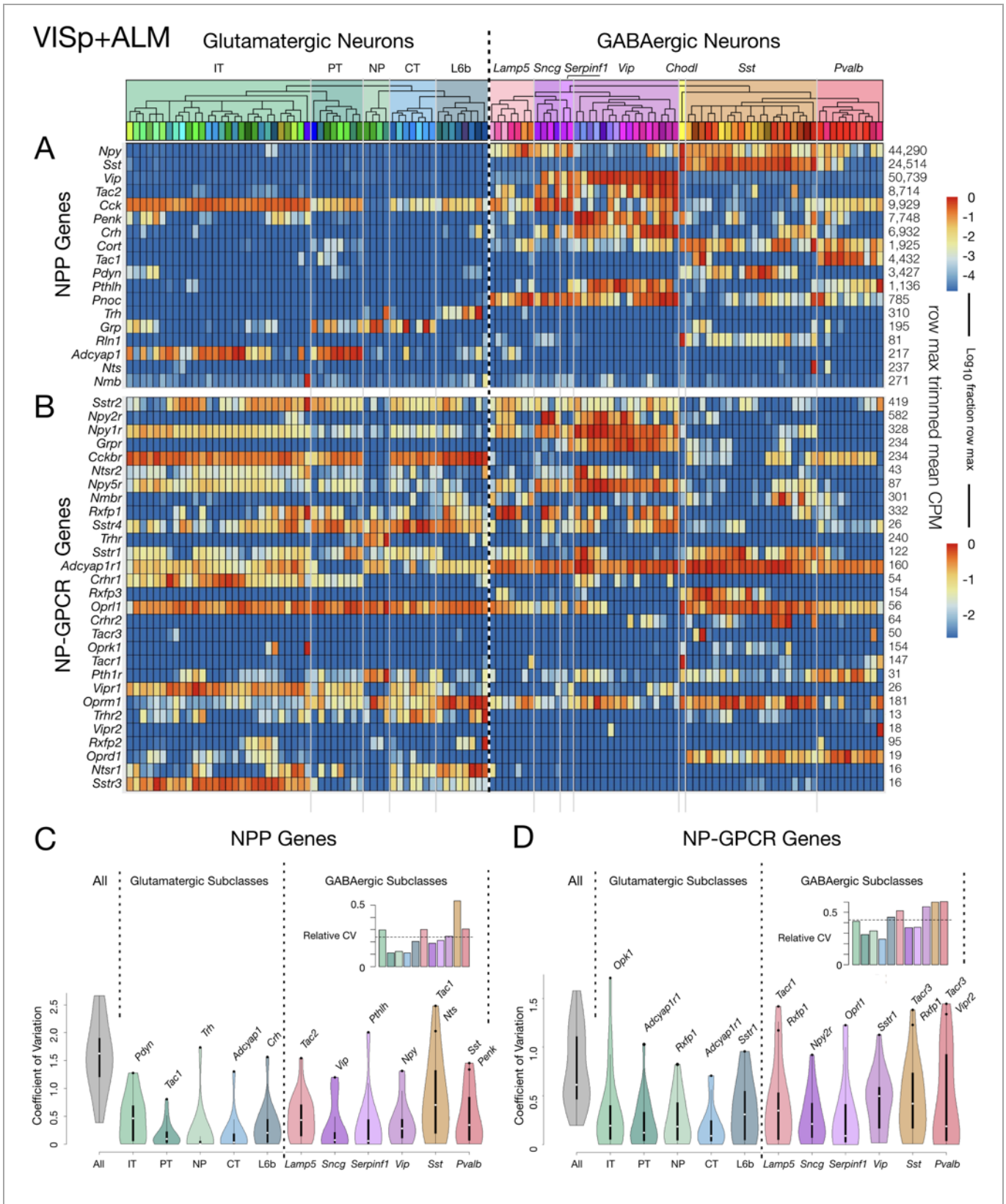


Figure 3. Neurotaxonomic heatmaps reveal highly neuron-type-specific expression of **(A)** 18 NPP and **(B)** 29 NP-GPCR genes in 22,439 individual neurons harvested from areas VISp and ALM. Trimmed-mean (5% trim) CPM expression values for each of the 115 VISp+ALM glutamatergic and GABAergic neuron types (see **Figure Supplement 1**) are normalized per gene to

maximum value indicated at right for each row and pseudocolored according to $\log_{10}$ scales at right. Note that these scales represent 5 (NPP) and 3 (NP-GPCR) orders of magnitude and that each gene spans the entire pseudocolor range across neuron types. Subclasses are called out here by labels (IT, PT, NP, CT, L6b for glutamatergic types; Lamp5, Sncg, Serpinf1, Vip, Chodl, Sst, Pvalb for GABAergic types) and demarcated on the heatmaps by thin gray lines. Gene rows are ordered here as in Tables 1 and 2. **(C)** Violin plots representing coefficients of CPM variation (CV) for 18 NPP genes across types pooled within each of the 11 subclasses indicated (Chodl not represented here as it is a singular neuron type) and globally across all cell types ("All"). Callouts on each violin indicate genes of highest CVs within each subclass. Inset shows within-subclass CV / global CV demonstrating variation within subclasses is a significant fraction of global variability (dotted line mean = 0.239). See **Figure supplement 2** for individual gene statistics. **(D)** Similar for 29 NP-GPCR genes showing greater relative variability, mean CV = 0.427. See **Figure supplement 3** for individual gene statistics.

The following source data and figure supplements are available for Figure 3:

Source data 1. Cell types, cluster ids and color codes for ALM and VISp regions.

Figure supplement 1. Tasic 2018 neurotaxonomy overview.

Figure supplement 2. NPP expression variation within subclasses.

Figure supplement 3. NP-GPCR expression variation within subclasses.

The dashed vertical line spanning **Figure 3A** and **3B** heatmaps divides glutamatergic and GABAergic neuron types and provides for ready comparison of NP gene expression patterns in these two broad neurotaxonomic classes. **Figure 3A** shows clearly that more NPP genes are expressed more strongly in GABAergic than in glutamatergic types. 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, Substance P), which has no parallel in the marking of glutamatergic neuron subsets. **Figure 3A** nonetheless also shows that every glutamatergic type expresses at least one NPP genes at a very substantial level. **Figure 3B** shows that the broader expression of NPP genes in GABAergic over glutamatergic types 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.

Additional graphics on the **Figure 3** heatmaps further delineate the Tasic 2018 neurotaxonomy. A cladogram reflects the hierarchical similarity progression from the broad GABAergic and glutamatergic classes to the 115 individual neuron types, as aggregated across VISp and ALM cortical areas. Tinted rectangles and labels call out the five glutamatergic and seven GABAergic subclasses (see also **Figure 3 - figure supplement 1)**. Thin gray vertical lines crossing both NPP and NP-GPCR heatmaps demarcate those same subclasses. This delineation of subclasses shows that expression of some genes tends to remain constant within some subclasses, but to change abruptly at subclass boundaries. This does not seem, however, to be a very general case. Many genes show expression that varies widely by type within subclass. **Figure 3C** quantifies such residual expression variation for all NPP and NP-GPCR genes within each subclass. These significant residuals justify the use of more narrowly defined taxa (e.g., the 115 neuron types) to adequately characterize cortical neuropeptide gene expression. Relationships between NP gene expression patterns and the Tasic 2018 neurotaxonomy will be examined more quantitatively in the following section.

Transcripts of 18 NPP and 29 NP-GPCR genes are exceptionally potent neuron-type markers

The strong marker patterning of the 47 NP gene expression profiles evident in **Figure 3** suggests the possibility that each of the 115 glutamatergic and GABAergic neuron types might be distinguished by a unique combination of these 18 NPP and 29 NP-GPCR genes. To explore this possibility and compare NP transcriptomes to other transcriptome subsets quantitatively, we developed the analysis presented in **Figure 4**.

We began 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 neuron's location in such a space can be inferred from the expression of a limited subset of genes (such as the 47 NP genes) would then provide a measure of sufficiency of that subset to determine

the type of that neuron accurately. Hierarchical clustering methods to define neuron types based upon gene expression are well established (Hastie et al., 2001; Oyelade et al., 2016) but have difficulty when comparing and making inferences between datasets. We therefore devised a machine learning approach based on linking deep neural networks called autoencoders (Hinton and Salakhutdinov, 2006) to address this question explicitly and quantitatively.

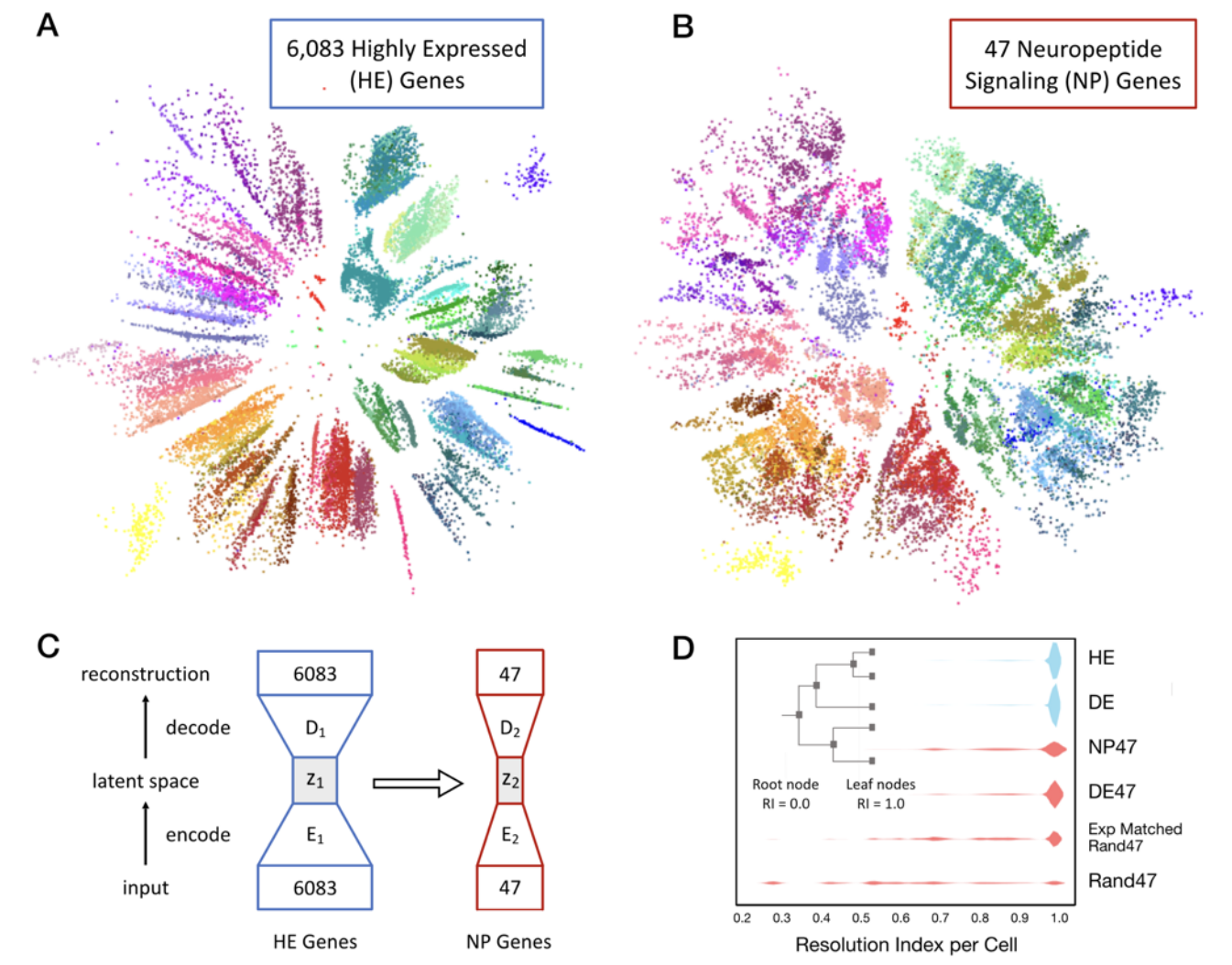


Figure 4. Neurons can be clustered effectively based on just 47 NP genes (18 NPP and 29 NP-GPCR). (A) A two-dimensional latent space representation of 22,439 cells based on 6,083 highly expressed (HE) genes obtained by an autoencoding neural network. Dots represent individual cells, colored according to the type-code assignments of Tasic et al. 2018 (see **Figure 3 - Figure supplement 1**). Cells of the same type appear as grouped into distinct islands, which suggests that classifiers trained to identify cell types would perform well with such low dimensional representations of gene expression. **(B)** Two-dimensional representation of neurons in such a latent space z_2, based on the 47 NP genes. **(C)** Schematic of the network architecture used to train the second autoencoder that learns to represent neurons in a latent space z_2 that is similar to z_1. This second autoencoder represents cells in the latent space based on much smaller gene sets. **(D)** Inset illustrates resolution index (RI) associated with nodes on the hierarchical tree used in the per-cell RI calculation. RI distribution (see Methods) for NP genes-based cell type classification shows that a vast majority of the cells can be correctly classified up to the type level (leaf nodes, RI=1.0) of the Tasic 2018 hierarchy. Errors in classification (RI<1.0) at the type level are nevertheless resolved at the class level of the hierarchy, as indicated by the high values for RI for the remaining cells. High average RI for HE genes, and 4,020 differentially expressed (DE) genes, and 47 DE genes indicates that the cell type classification procedure based on autoencoder representations is accurate. The average RI for cell type classification based on the 47 NP genes is significantly higher ($p<0.01$, bootstrap) than both, subsets of 47 genes selected randomly (Rand47, n=100 subsets), and selected randomly but with expression levels matched to the NP genes (Rand47 ExpMatch, n=100 subsets).

The following figure supplement is available for figure 4:

Figure supplement 1. Autoencoder expression clustering performance by neuron type.

Figure supplement 2. Determination of optimal autoencoder latent space dimensionality.

We trained a single autoencoder network to represent cells in a low dimensional space based on CPM values of the 6,083 most highly expressed genes (HE genes) in the Tasic 2018 dataset. **Figure 4A** shows the result of one such two dimensional encoding, where each of the 22,439 individual neurons appear as a distinct dot colored by its type assignment. The tight grouping of type-coding colors evident in **Figure 4A** implicitly conveys that position within this latent space corresponds well to neuron types, despite the fact that the autoencoder did not have prior information about the Tasic 2018 classification. With the first autoencoder held as fixed, we trained a second autoencoder, linked to the first, to obtain a low-dimensional representation based on a much smaller subset of genes. **Figure 4B** shows a two-dimensional representation of the same 22,439 neurons based on 47 NP genes. Again the tight color grouping suggests that the 47 NP genes alone suffice to assign types in close register to the Tasic 2018 neurotaxonomy. The autoencoder network architectures are schematized in **Figure 4C**. The cost function used to train the second autoencoder included a penalty term to minimize differences in the representation of cells compared to that obtained by the first autoencoder. This was done to ensure that the latent spaces of the two autoencoders are as similar as possible while faithfully representing the expression patterns of the respective gene sets they receive as input. This procedure allowed us to visualize the similarity between the gene sets in a latent space that captures type information, and to quantify the extent to which any small gene subset by itself could be used to identify neuron types.

To quantify the type classification ability of different gene sets, we used Quadratic Discriminant Analysis (QDA) (Hastie et al., 2001) to perform supervised classification using five-dimensional latent space representations of the different gene sets obtained by autoencoder networks. We obtained a measure on a per-cell basis, resolution index (RI) to evaluate the degree of correspondence of classification results with the Tasic 2018 neurotaxonomy. The resolution index averaged over all cells is used as a summary statistic to quantify the ability of different gene sets in resolving neuron types. Briefly, QDA was performed iteratively on a given latent space representation, starting with all the leaf node type labels of the neurotaxonomy. In each subsequent iteration the number of labels was reduced by successively merging leaf node labels into their parent node class label (inset, **Figure 4D**). RI=1.0 for a neuron that is assigned the correct type (e.g., Pvalb Reln Tac1) and 0.0<RI<1.0 for neurons for which the iterative QDA based classification could determine the correct label only up to a subclass (e.g. Pvalb). A neuron is assigned RI=0.0 if the QDA based classification failed to determine the correct label even at the glutamatergic or GABAergic level.

Figure 4D shows neuron type classification results based on five dimensional latent space representations of different subsets of genes (k=13 fold cross validation). For the 6,083 HE genes and a set of 4,020 genes most differentially expressed (DE genes) across neuron types, the latent space is obtained with the first autoencoder, and the RI distributions shown in blue have average values of 0.986 and 0.987 respectively, close to the theoretical maximum of 1.0 that can only be achieved with perfect type classification for all neurons in the dataset. For subsets of 47 genes, the latent representations were obtained with the second linked autoencoder, and the corresponding RI distributions are colored red. A set of 47 DE genes achieves average RI=0.964. These results confirm the idea that autoencoder-based low dimensional representations of gene expression can be used for accurate type classification. The 47 NP genes can be used to classify neuron types well, with average RI=0.925 and a majority of the neurons (62%) classified correctly at the type level (with nearly uniform performance across all neuron types, see **Figure 4 – Supplemental figure 1**). This RI performance is significantly higher ($p<0.01$, bootstrap) than the average RI for of subsets of genes chosen randomly (0.641±0.047, n=100), and chosen randomly but with expression levels matched with the NP genes (0.843±0.027, n=100), with none of the individual randomly selected subsets reaching the NP gene index of 0.925. Note that genes in the 47 DE set were chosen with prior knowledge of the Tasic 2018 taxonomy, while the 47 NP gene set was not. This distinction thus makes the near match of the 47 NP to the 47 DE gene sets in average RI all the more striking. This demonstration of the exceptional power of NP genes to mark transcriptomic neuron types

reinforces earlier indications of an especially close and fundamental connection between neuropeptide gene expression and neuron type identity.

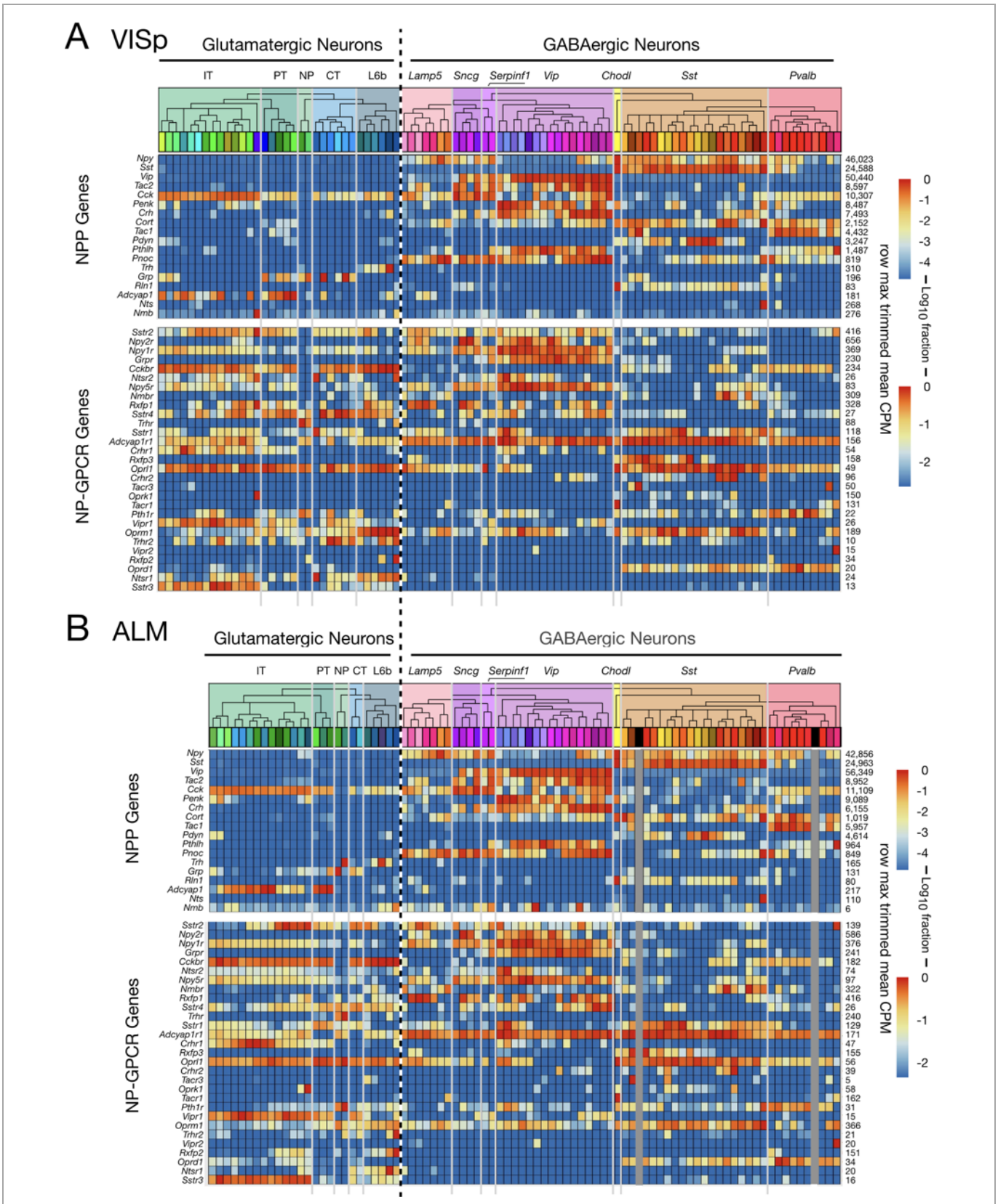


Figure 5. Neurotaxonomic heatmaps show strong conservation of NPP and NP-GPCR expression patterns between two distant neocortical areas. (A) Expression heatmap for 18 NPP and 29 NP-GPCR genes in 13,491 single VISp neurons classified by type. **(B)** Similar heatmap for 8,948 single neurons harvested from ALM. Heatmaps generated and displayed as described

in Figure 3, except for segregation here of VISp and ALM harvest areas. Heat maps are aligned horizontally here to match GABAergic neuron types between VISp and ALM. Vertical dark gray bars in figure 5B are spacers marking the two GABAergic cell types absent in ALM. Glutamatergic neurotaxonomies are seen to differ substantially, but differences appear mainly at the finest, "leaf" levels of the neurotaxonomic hierarchy (see Tasic 2018 and **Figure 3 - Figure supplement 1**).

The follow source data are available for figure 5:

Source data 1. Cell types, cluster ids and color codes for VISp region.

Source data 2. Cell types, cluster ids and color codes for ALM region.

Conservation of NPP and NP-GPCR gene expression profiles between VISp and ALM

Figure 5 juxtaposes separate VISp and ALM expression profiles for NPP and NP-GPCR genes across 93 VISp neuron types (**Figure 5A**) and 84 ALM neuron types (**Figure 5B**). Similarities of expression profiles for the two areas are obvious in **Figure 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., 2018. 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 dark gray vertical placeholder bars in **Figure 5B**). The general conservation of neuron-type-specific expression patterns among common cell types 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 another indication of robust connection between NP gene expression and cortical neuron differentiation.

Prediction of local peptidergic signaling from expression of cognate NPP/NP-GPCR pairs

Expression of an NPP gene in one neuron and a cognate NP-GPCR gene in another neuron nearby implies a possibility of directed paracrine signaling, with diffusion of a secreted peptide coupling the first neuron to the second. 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 predicts accordingly 37 distinct peptidergic neuromodulation 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 therefore make predictions of 74 (37 x 2) peptidergic distinct signaling networks, keeping separate consideration of signaling within VISp and within ALM.

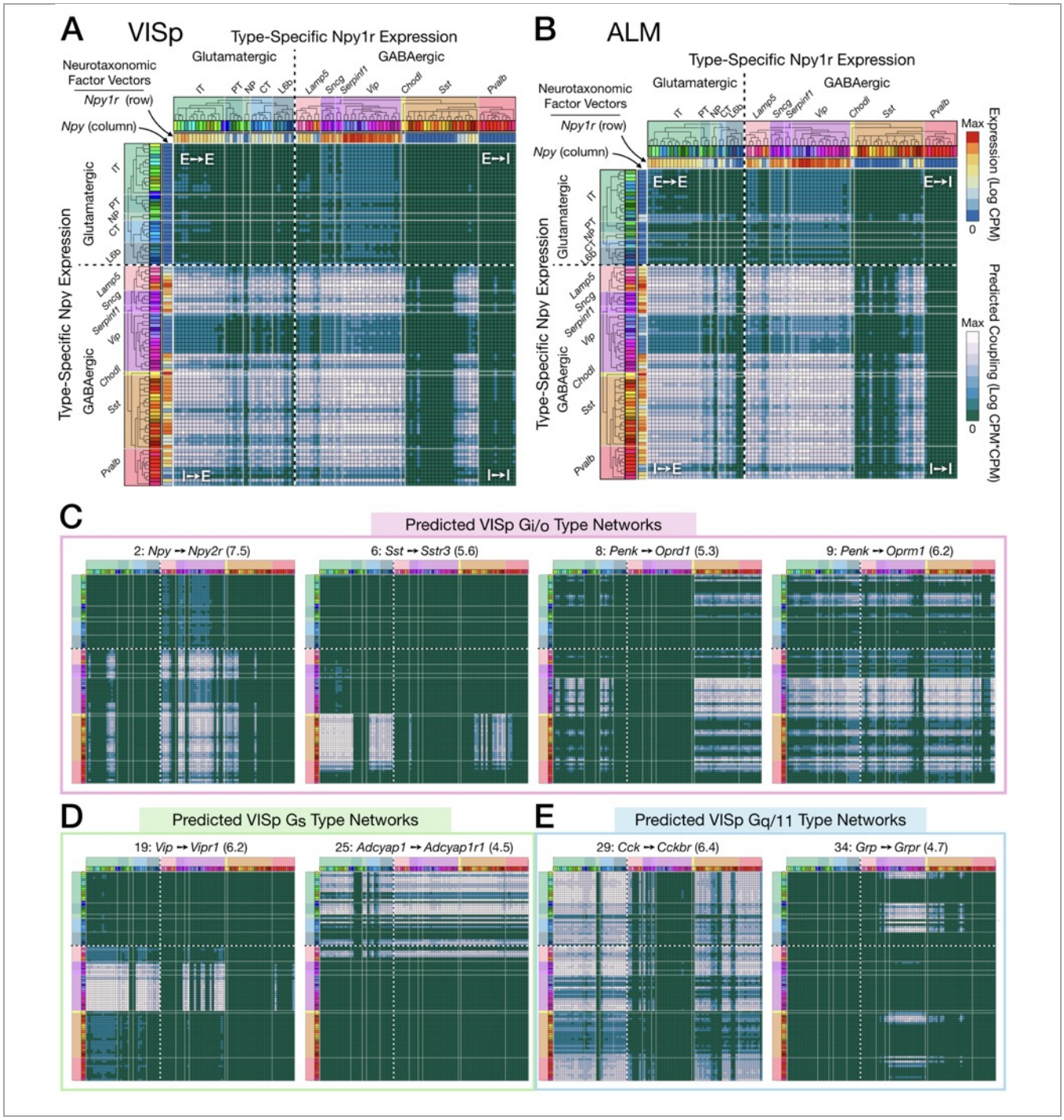
#	Cognate Pair Symbol	NPP Gene	NP-GPCR Gene	Primary G_α Family	Fraction of Type Pairs
1	*Npy→Npy1r*	*Npy*	*Npy1r*	$G_{i/o}$	0.7805
2	*Npy→Npy2r*	*Npy*	*Npy2r*	$G_{i/o}$	0.341
3	*Npy→Npy5r*	*Npy*	*Npy5r*	$G_{i/o}$	0.8095
4	*Sst→Sstr1*	*Sst*	*Sstr1*	$G_{i/o}$	0.751
5	*Sst→Sstr2*	*Sst*	*Sstr2*	$G_{i/o}$	0.836
6	*Sst→Sstr3*	*Sst*	*Sstr3*	$G_{i/o}$	0.405
7	*Sst→Sstr4*	*Sst*	*Sstr4*	$G_{i/o}$	0.806
8	*Penk→Oprd1*	*Penk*	*Oprd1*	$G_{i/o}$	0.4955
9	*Penk→Oprm1*	*Penk*	*Oprm1*	$G_{i/o}$	0.9
10	*Cort→Sstr1*	*Cort*	*Sstr1*	$G_{i/o}$	0.6265
11	*Cort→Sstr2*	*Cort*	*Sstr2*	$G_{i/o}$	0.6965
12	*Cort→Sstr3*	*Cort*	*Sstr3*	$G_{i/o}$	0.338
13	*Cort→Sstr4*	*Cort*	*Sstr4*	$G_{i/o}$	0.672
14	*Pdyn→Oprd1*	*Pdyn*	*Oprd1*	$G_{i/o}$	0.2115
15	*Pdyn→Oprk1*	*Pdyn*	*Oprk1*	$G_{i/o}$	0.0745
16	*Pdyn→Oprm1*	*Pdyn*	*Oprm1*	$G_{i/o}$	0.4
17	*Pnoc→Oprl1*	*Pnoc*	*Oprl1*	$G_{i/o}$	0.654
18	*Rln1→Rxfp3*	*Rln1*	*Rxfp3*	$G_{i/o}$	0.106
19	*Vip→Vipr1*	*Vip*	*Vipr1*	G_s	0.496
20	*Vip→Vipr2*	*Vip*	*Vipr2*	G_s	0.052
21	*Crh→Crhr1*	*Crh*	*Crhr1*	G_s	0.3925
22	*Crh→Crhr2*	*Crh*	*Crhr2*	G_s	0.2035
23	*Rln1→Rxfp1*	*Rln1*	*Rxfp1*	G_s	0.2465
24	*Rln1→Rxfp2*	*Rln1*	*Rxfp2*	G_s	0.07
25	*Adcyap1→Adcyap1r1*	*Adcyap1*	*Adcyap1r1*	G_s	0.284
26	*Adcyap1→Vipr1*	*Adcyap1*	*Vipr1*	G_s	0.1465
27	*Adcyap1→Vipr2*	*Adcyap1*	*Vipr2*	G_s	0.0155
28	*Tac2→Tacr3*	*Tac2*	*Tacr3*	$G_{q/11}$	0.0955
29	*Cck→Cckbr*	*Cck*	*Cckbr*	$G_{q/11}$	0.6635
30	*Tac1→Tacr1*	*Tac1*	*Tacr1*	$G_{q/11}$	0.119
31	*Pthlh→Pth1r*	*Pthlh*	*Pth1r*	$G_{q/11}$	0.392
32	*Trh→Trhr*	*Trh*	*Trhr*	$G_{q/11}$	0.016
33	*Trh→Trhr2*	*Trh*	*Trhr2*	$G_{q/11}$	0.055
34	*Grp→Grpr*	*Grp*	*Grpr*	$G_{q/11}$	0.113
35	*Nts→Ntsr1*	*Nts*	*Ntsr1*	$G_{q/11}$	0.0225
36	*Nts→Ntsr2*	*Nts*	*Ntsr2*	$G_{q/11}$	0.054
37	*Nmb→Nmbr*	*Nmb*	*Nmbr*	$G_{q/11}$	0.5655

Table 3. The 18 NPP and 29 NP-GPCR genes of Tables 1 and 2 constitute 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 (Alexander et al., 2017) and a fraction denoting

frequency with which the given cognate pair occurs as a fraction of all neuron pairs surveyed. Pastel table fill colors denote G-protein transduction class as in Tables 1 and 2.

Prediction of peptidergic networks from neurotaxonomic NP gene expression profiles

Figure 6 displays weighted adjacency matrix plots representing predictions of neuron-type-specific and neuron-subclass-specific peptidergic coupling from selections drawn from VISp and ALM of the 37 cognate NP gene pairs. The prediction matrices A-E are outer products (CPM*CPM units) of vectors representing expression (CPM units) of an NPP gene (columns) and a cognate NP-GPCR gene (rows) across all VISp or ALM neuron types. The predicted coupling matrices in F matrices are similar except that factor vectors are down-sampled by averaging neuron-type-specific CPM values within each of the subclasses (see **Materials and Methods** for more details).



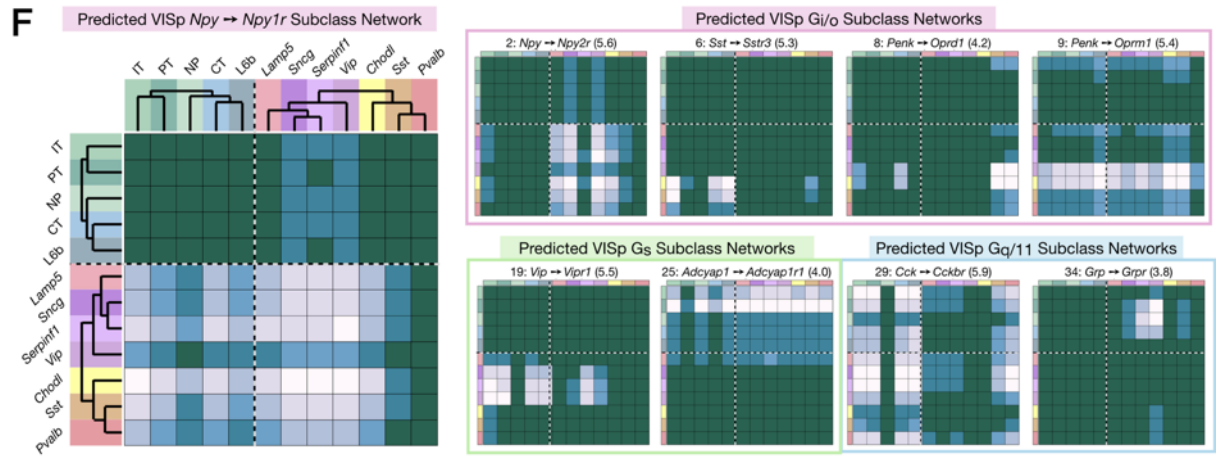


Figure 6. Neurotaxonomic expression profiles of 37 cognate NPP/NP-GPCR pairs predict 37 peptidergic networks. Weighted adjacency matrix plots predicting local peptidergic coupling amongst neuron types **(A-E)** and subclasses **(F)**. Matrices were computed as outer products (CPM*CPM units) of row and column factor vectors representing abundance (CPM units) of NPP and cognate NP-GPCR genes. Pseudocolor scales representing both expression (CPM) and coupling (CPM*CPM) are logarithmic. **(A)** 93 x 93 matrix predicting coupling amongst 93 VISp neuron types based on type-specific expression of the Npy gene and the cognate Npy1r NP-GPCR gene, as indicated by row and column vector "heat" strips called out by curved arrows. **(B)** An 84 x 84 square matrix similarly representing Npy-Npy1r coupling, depicted as in (A), except based only on the 84 ALM neuron types. Dashed crosses demarcate the four quadrants of directed NPP/NP-GPCR pairing between glutamatergic ("E") and GABAergic ("I") neuron types, called out as (E→ E), (E→ I), (I→ I) and (I→ E). Light gray lines, pastel color blocks and labels flanking both axes demarcate higher, subclass levels of the Tasic 2018 neurotaxonomy (as in Figure 3A above). **(C-E)** Exemplar matrix predictions for a further sampling of the 37 VISp cognate NPP/NP-GPCR pairs from each of the three primary G-protein transduction families: (C) Gi/o; (D) Gs; (E) Gq/11. **(F)** Adjacency matrices similar to A-E, except row and column factor vectors were calculated as means of CPM values across neuron types comprising indicated subclasses. (Cladograms and taxonomic color codes as delineated in **Figure 3 - Figure supplements 1A-H**). Links below point to source data files and similar plots for all 37 VISp and ALM type and subclass adjacency matrices, and to additional quantitative analysis of coupling matrix hierarchies (**Figure 6 – Figure supplement 2**) and morphologies and correlations (**Figure 6 – Figure supplement 3**).

The following source data and figure supplements are available for figure 6:

Source data 1. NP coupling predictions by cognate pair and type for area VISp (archive containing 37 CSV files, one for each of the 37 cognate pairs listed in Table 3 and represented in Figure supplements 1A-H).

Source data 2. NP coupling predictions by cognate pair and type for area ALM (archive containing 37 CSV files, one for each of the 37 cognate pairs listed in Table 3 and represented in Figure supplements 1A-H).

Figure supplements 1A-H Network coupling predictions by type and subclass for 37 cognate NPP/NP-GPCR pairs in VISp and ALM.

Figure supplement 2. Coupling matrix hierarchies.

Figure supplement 3. Coupling matrix localization and correlation metrics.

Figure 6C-E represent 8 more of the 37 cognate pair coupling matrices predicted for VISp. Along with Figures **6A** and **6B**, these exemplify the wide variety of neuron-type-specific coupling motifs resulting from transcriptomic prediction. Most coupling matrices (i.e., pairs 1, 9, 29), predict significant coupling over wide swaths of type-pairs, approaching 20% of the entire matrix. A few matrices at the other extreme, such as 6 and 34, predict very sparse coupling. Other predictions are intermediate in sparsity. As one might expect, similar patterns are evident in the downsampled, subclass level predictions of **Figure 6F**. Even from the small subset of the 37 coupling matrix plots shown in **Figure 6**, it is evident that both type-level and subclass-level matrices are densely tiled by predictions of connectivity. Inspection of **Figure 6** and similar plots for the remainder of the 37 cognate pairs (**Figure 6 – Supplemental figures 1A-1H**) also reveals that there is a great deal of cross-network redundancy, with multiple pairs covering a

large majority of the coupled types and subclasses, sometimes within and sometimes crossing G_α family boundaries. These observations will be strengthened by the analysis of **Figure 7** below.

Finally, **Figure 6** illustrates the tendency of coupling predictions from most 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 Tasic 2018 hierarchical clustering which was also evident in **Figures 3** and **5**. The broadest example of coupling matrix patches reflecting hierarchical neurotaxonomy structure is provided by the observation that most sizable coupling patches fall strictly within single quadrants of glutamatergic-GABAergic neuron type pairing. Variations in coupling matrix structure across all 37 cognate NP pairs are represented in more quatitative terms by **Figure 6 - Supplemental figures 9** and **10**. Additional details regarding the generation of the coupling matrices are provided in **Materials and Methods.**

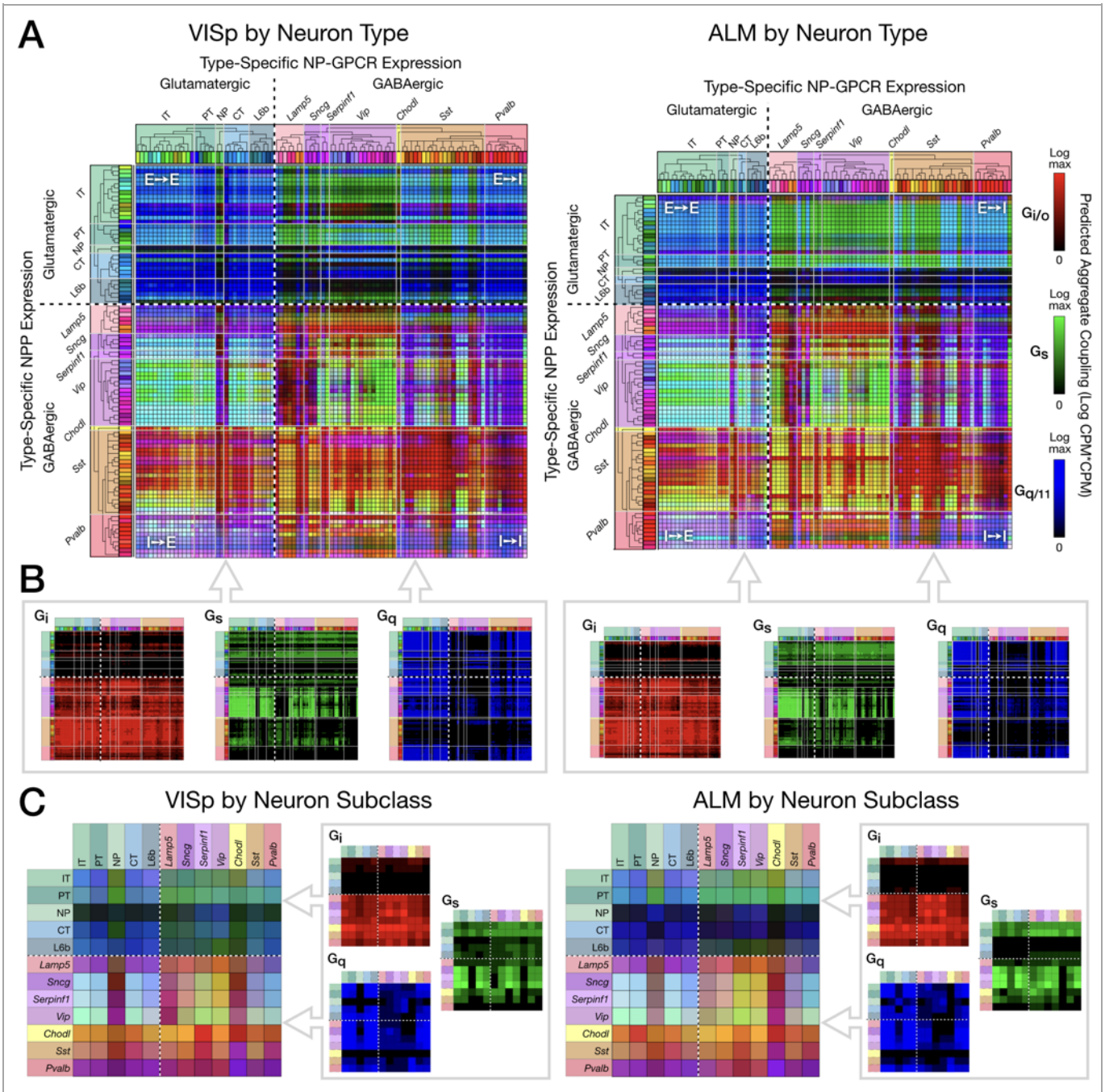


Figure 7. Pooling NP-GPCRs by primary G_α family enables neurotaxonomic prediction of primary NP-GPCR signaling impacts across 37 cognate NP pairs (see text) for each of areas VISp and ALM. **(A)** ISQ color maps representing coupling matrix

predictions at the Tasic 2018 neurotaxonomy type level, merging Gi/o (red), Gs (green) and Gq/11 (blue) primary G_α family components. **(B)** Component primary G_α family (color) channels prior to merger displayed in (A). **(C)** Coupling matrix plots as in (A) and (B), except generated at the higher Tasic 2018 neurotaxonomy subclass level. Individual aggregate matrix components as in (B) are plotted at the right for both VISp and ALM. Dashed white crosses overlaying each matrix plot demarcate glutamatergic and GABAergic classes and the four corresponding matrix quadrants (E→ E, E→I, I→I and I→E) as in the individual matrix plots of **Figure 6**. (Cladograms and taxonomic color codes as delineated in **Figure 3 - Figure supplement 1**).

The following source data are available for figure 7:

Source data 1. VISp coupling predictions by Gα family and type. (archive containing three CSV files, one for each of Gi/o, Gs and Gq/11 families).

Source data 2. ALM coupling predictions by Gα family and type. (archive containing three CSV files, one for each of Gi/o, Gs and Gq/11 families).

Prediction of second-messenger impacts from neurotaxonomic NP gene expression profiles

For compact visualization of predicted signaling impacts of multiple distinct peptidergic networks and to facilitate empirical tests of such predictions based on calcium and cyclic AMP sensors (see **Discussion**), we developed the "ISQ" graphic exemplified in **Figure 7**. This treatment makes use of the trichotomous G-protein primary transduction family approximation described in **Introduction** and delineated in **Table 3** above: the $G_{i/o}$ family ("I") inhibits production of cyclic AMP, the G_s family ("S") stimulates production of cyclic AMP) and the $G_{q/11}$ family ("Q") augments intracellular calcium dynamics. Such trichotomy is certainly an oversimplification, as it is known that downstream GPCR signal transduction is richly multifacted (Weis and Kobilka, 2018) and that some GPCRs may signal via members of multiple G_α families, but we postulate here that this simplified scheme may nonetheless offer a first approximation useful for the design of exploratory experimentation and theory.

Figure 7 displays three-channel "ISQ" (red, green, blue) color maps predicting coupling in areas VISp and ALM based on aggregation across three primary G_α families ($G_{i/o}$, G_s and $G_{q/11}$). Individual cognate-pair coupling matrices were computed as in **Figure 6**, log10 scaled, individually normalized to maximum values, then summed into a red, green or blue color channel by primary G_α family as listed in **Table 3**. **Figure 7A** merges red, green and blue (i.e., $G_{i/o}$, G_s, $G_{q/11}$) color channels for the Tasic 2018 neuron-type-level. **Figure 7B** displays the three component channels individually. The dashed white crosses on these and following coupling matrix both plots divide these ISQ maps into four E-I quadrants as in **Figure 6**. Major features of the ISQ maps are clearly very similar for VISp and ALM. **Figure 7**C show aggregated matrix plots generated in similar fashion for the subclass-level neurotaxonomy.

The ISQ maps of **Figure 7** exhibit a number of interesting features. (1) The aggregate matrices show that the 37 cognate pairs cumulatively predict coupling that densely tiles the entire neuron-type coupling matrix, with the largest area of relatively weak coupling being that from NP, CT and L6b subclasses to GABAergic neurons. (2) Aggregate predictions are highly conserved between VISp and ALM areas. (3) All four E-I quadrants show coupling representative of all three G_α families. (4) There is nonetheless some family predominance within each quadrant: $G_{i/o}$ (blue) in the E→E quadrant, G_s (green) in the E→I quadrant, and $G_{i/o}$ (red) in the I→I and I→E quadrants. (5) As is particularly notable in the component matrix plots, $G_{i/o}$ (red) signaling is the most heavily concentrated, with quite little signaling expressed in the top two (E-E and E-I) quadrants but tiling the bottom two (I-E and I-I) quite thoroughly. $G_{q/11}$ signaling shows a weaker, but still noticeable tendency toward concentration in the two left quadrants (E-E and I-E). G_s signaling exhibits distinct zone of concentration, but these are not well captured by the quadrant structure. (6) The presence of cyan, yellow and purple (blended) colors in the merged matrix plots (A and C), particularly in the bottom quadrants is indicative of coincidence of signaling impacts of multiple G_α families at individual type (A) and subclass (C) intersections.

Discussion

Light from single-cell transcriptomics is now beginning to illuminate dark corners of cellular neuroscience that have long resisted mechanistic and functional analysis (Fan et al., 2018; Fishell and Kepecs, 2019; Földy et al., 2016; Gokce et al., 2016; Luo et al., 2018; Okaty et al., 2011; Paul et al., 2017; Shekhar et al., 2016; Tasic et al., 2018, 2016; Telley et al., 2016; Zeng and Sanes, 2017). 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 (Borbély et al., 2013; Burbach, 2011; Elphick et al., 2018; Grimmelikhuijzen and Hauser, 2012; Katz and Lillvis, 2014; Kuffler et al., 1979) and there certainly is an excellent literature on cortical neuropeptide signaling (Crawley, 1985; Férézou et al., 2007; Gallopin et al., 2006; Gomtsian et al., 2018; Hamilton et al., 2013; Liu et al., 2018; Mena et al., 2013; Mitre et al., 2018; Owen et al., 2013; Rossier and Chapouthier, 1982; Williams and Zieglgänsberger, 1981), 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 relative rarity. Though many NPP and 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 often redundant. These previously uncharted differential expression factors may have hindered repeatable experimentation. Our analysis supports this unwelcome proposition but may also point the way to more productive new perspectives on intracortical peptidergic neuromodulation.

Summary of findings

The present single-cell analysis establishes that mRNA transcripts from one or more of 18 NPP genes are detectible in over 97% of mouse neocortical neurons (**Figure 1A,B**) and that transcripts of one or more of 29 cognate NP-GPCR genes are detectible in over 98% (**Figure 2A,B**). Transcripts of at least one of the 18 NPP genes are present in the vast majority of cortical neurons at extremely high copy number (**Table 1**), suggesting the likelihood of brisk translation into neuropeptide precursor proteins. Brisk synthesis of precursor proteins further suggests brisk processing to active neuropeptide products and secretion of these products. Likewise, NP-GPCR transcripts rank high in abundance compared to most other transcripts of protein-coding genes (**Table 2**), supporting the likelihood of functional receptor products. Our observations thus support the proposition that all, or very nearly all, neocortical neurons, both glutamatergic and GABAergic, are also both neuropeptidergic and modulated by neuropeptides. We are not aware of any previous empirical support for quite such a strong conclusion.

Leveraging the analytical power of the Tasic 2018 transcriptomic neurotaxonomy, we find that patterns of differential expression of the 18 NPP and 29 NP-GPCR genes are very highly specific to neuron types as discerned from genome-wide transcriptomic analysis (**Figure 3**). Though much additional work (e.g., see Cadwell et al., 2017; Daigle et al., 2018; Moffitt et al., 2016; Shah et al., 2016; Wang et al., 2018; Zeng and Sanes, 2017) will be needed to fully reconcile new transcriptomic neurotaxonomies such as the Tasic 2018 example with existing anatomical and physiological neurotaxonomies, it seems very likely that some such reconciliation will eventually take place, and that the dimensions of neurotaxonomy will be expanded to include emerging connectomic data (Jonas and Kording, 2015).

Our analysis shows that very intricate single-cell (**Figures 1B,C** and **2B,C**) and neurotaxonomic (**Figure 5**) patterns of expression of 18 NPP and 29 cognate NP-GPCR genes are very rigorously conserved between VISp and ALM, two distant and quite different areas of neocortex. Such strong conservation would seem improbable if these intricate patterns resulted from ephemeral factors such as local electrical activity or modulation status. Rather, we suggest that this strong conservation is more likely to reflect a really fundamental evolutionary and developmental connection between neuropeptide network architectures and adaptive cortical circuit function.

Following earlier indications that neurons may express multiple NPP genes, e.g., (Mezey et al., 1999), our analysis establishes that expression of multiple NPP genes in individual neurons may be the rule in cortex

(**Figure 1C**). Our analysis also establishes the generality of expression of multiple NP-GPCR genes in individual cortical neurons (**Figure 2C**). The significance of these observations remains to be explored but should be viewed in light of recent discoveries of large numbers and great diversity of transcriptomic neuron types in neocortex and many other brain regions. Combinatorial expression of neuropeptide precursor and receptor genes obviously expands the prospects for molecular multiplexing that may allow selective communication amongst a multiplicity of distinct neuron types even though the signaling molecules propagate in diffuse paracrine fashion. It is also good to keep in mind, however, that the selectivity of NP-GPCRs for particular peptide moieties is not perfect. Various kinds of concentration-dependent "crosstalk" between nominally separate peptidergic networks are therefore possible. Here in the interests of simplicity we have confined explicit peptidergic signaling predictions to the highest affinity pairings of NPP and NP-GPCR gene products (e.g., see Alexander et al., 2017).

We also find that a modest set of 47 neuropeptide-signaling genes permits transcriptomic neuron type classification that is exceptionally precise in comparison to other similarly small gene sets (**Figure 4**). Connections between neuronal cell-type differentiation and differential expression of neuropeptides were first recognized by the widespread use of neuropeptide immunoreactivity to discriminate interneuron types (DeFelipe et al., 2013). The exceptional power of neuropeptide genes as cell type markers is also evident in the Tasic 2018 neuron-type nomenclature (see (Tasic et al., 2018) and bold red type highlights in **Figure 3 – figure supplement 1**) and is noteworthy in other recent single-cell transcriptomic analyses of mouse neuron differentiation (Huang and Paul, 2019; Paul et al., 2017; Sugino et al., 2019; Zeisel et al., 2018). The tight alignment of neuron type classifications based solely on neuropeptide-signaling gene expression with the classifications based on genome-wide expression patterns, as evident in **Figure 4**, offers an intriguing suggestion of a very deep and fundamental connection between the expression of evolutionarily ancient neuropeptide-signaling genes and the differentiation of neuron type identities during metazoan speciation.

The structures of predicted neuropeptidergic modulation networks

Our analysis delineates neuron-type-specific expression of 37 cognate pairs amongst the 18 NPP and 29 NP-GPCR genes analyzed (**Table 3**). Each of these pairs can be taken to predict a modulatory connection from cells expressing a particular NPP gene, via a secreted NP product, to cells expressing the particular NP-GPCR gene (**Figure 6**). Each pair thus establishes the prospect of a directed modulatory network with nodes defined by the neurotaxonomic identities of the transmitting NPP-expressing and the receiving NP-GPCR-expressing neurons. The analyses represented in **Figures 1, 2, 3** and **5** and **Table 3** establish that at least one of the 37 pairs directly involves every neuron sampled, and that the vast majority of neurons are directly involved in more than one of the 37 predicted networks. The nearly complete adjacency matrix tiling evident in **Figures 6** and **7** remarkably suggests that at least one of the 18 peptides considered here may directly interconnect almost every cortical neuron type with almost every other neuron type. Because of this saturated, multiplexed coverage of all neurons and neuron types, we refer to these predicted neuropeptidergic networks as "dense".

Transcriptomic prediction of paracrine 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 secreted neuropeptides act locally. On the other hand, most of the glutamatergic neurons do emit long axons, so it is possible that neuropeptides secreted from such neurons may act in remote cortical or extracortical projection target areas. Even so, most cortical glutamatergic neurons also have locally ramifying axon branches and may also secrete neuropeptides from their local dendritic arbors (Vila-Porcile et al., 2009). The high cortical expression of NP-GPCRs cognate to NPP genes expressed by glutamatergic neurons in the same local area suggests a scenario supportive of local modulatory signaling from glutamatergic neuron sources, though this case may not be quite as strong as that for strictly local GABAergic neurons. That said, the much more profuse expression of NPP genes in

GABAergic neuron types along with the somewhat more profuse NP-GPCR expression in glutamatergic types does suggest a "prevailing wind" of peptidergic signaling, blowing mainly from GABAergic to glutamatergic neurons, as presaged in an earlier microarray analysis of developing mouse cortex (Batista-Brito et al., 2008).

Though our NP network predictions are entirely consistent with decades of pioneering work on peptidergic neuromodulation and cortical gene expression (Burbach, 2011; Hökfelt et al., 2013; van den Pol, 2012), it is only with the recent advent of single-cell and neurotaxonomics methods that such specific predictions have become possible and, most importantly, testable.

Testing peptidergic network predictions

The present predictions regrading cortical neuropeptidergic coupling are based on detection of cellular mRNA transcripts, but prediction from such data depends upon (1) extrapolation from cellular mRNA census to inference about rates of 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 signaling consequences of neuropeptide binding to cortical NP-GPCR receptors. Though we have already discussed several factors that mitigate such concerns, we stipulate here that these uncertainties remain substantial. Nonetheless, we expact that these same uncertainties will define paths for very productive future research.

Physiological and anatomical experimentation will be essential to testing transcriptomic predictions of intracortical neuropeptide signaling. We have suggested that such work may have been frustrated in the past by irreproducibility due to the uncharted multiplicity, neuron-type-specificity, and redundancy of NPP and NP-GPCR expression. This conundrum may now be resolved with the emergence of transcriptomic neurotaxonomies and new tools for experimental access to specific cortical neuron types. Such access may be either ***prospective***, using Cre and/or Flp driver lines (Daigle et al., 2018; He et al., 2016; Madisen et al., 2015) or viral vectors (Dimidschstein et al., 2016) of substantial neuron-type-specificity, or ***retrospective*** by multiplexed FISH (Lein et al., 2017; Zeng and Sanes, 2017), immunostaining (He et al., 2016; Xu et al., 2010), patch-seq (Cadwell et al., 2017; Lein et al., 2017) or morphological classification methods (DeFelipe et al., 2013; Zeng and Sanes, 2017). These and other new molecular tools like those discussed below now seem poised enable truly decisive and repeatable tests of neuron-type-specific transcriptomic predictions of peptidergic signaling. It will be critical, however, for the field to have continually updated access to rapidly growing bodies of genetic and transcriptomic data and to the requisite animal strains and labeling materials.

A vast pharmacopeia of well-characterized specific ligands and antagonists for most NP-GPCRs (Alexander et al., 2017) will be 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 foundations for measurement of neuropeptide impacts. In addition, many new tools more specific to neuropeptide signaling are emerging. Super-resolution 3D immunohistologies like array tomography (Smith, 2018) and 3D single-molecule methods (Jia et al., 2014; von Diezmann et al., 2017) will enable imaging of dense-core vesicle localization and neuropeptide contents in type-specific network anatomical context. Genetically encoded fluorescent dense-core vesicle cargos will allow real-time detection of neuropeptide secretion (Ding et al., 2019), while genetically encoded sensors of extracellular GPCR ligands (Patriarchi et al., 2018; Sun et al., 2018), GPCR activation (Haider et al., 2019; Hill and Watson, 2018; Livingston et al., 2018; Ratnayake et al., 2017; Stoeber et al., 2018), G-protein mobilization (Ratnayake et al., 2017), cAMP concentration (Hackley et al., 2018; Ma et al., 2018), protein kinase activation (Chen et al., 2014) and protein phosphorylation (Haider et al., 2019) will enable fine dissection of NP dynamics and NP-GPCR signal transduction events (Spangler and Bruchas, 2017). In addition, new caged NP-GPCR ligands (Banghart et al., 2018) and antagonists (Banghart et al., 2013) will provide for precise spatial and temporal control for NP receptor activation. All

of these tools have already been proved at least in principle, and all should be readily applicable to testing specific hypotheses derived from the type-specific peptidergic signaling predictions set forth here (**Figures 6, 7 and their supplements**).

Prospects for elucidating cortical homeostasis, modulation and plasticity

Our results suggest that densely multiplexed peptidergic networks could play very significant roles in the homeostasis, modulation and plasticity of cortical synaptic networks. Due to the clearly formidable complexity of cortical networks, however, a real grasp of the myriad network interactions implicated is certain to require theoretical and computational approaches, in addition to experimental biophysics tests as outlined in the preceding section. Work at the fertile intersection of the neuroscience and the computer science of learning (Dayan and Abbott, 2001; Huh and Sejnowski, 2017; Koch and Segev, 1998; Lillicrap et al., 2016; Marblestone et al., 2016; Shai and Larkum, 2017; Song et al., 2000) seems particularly relevant to fathoming the possible significance of the neuropeptidergic networks we predict here.

Neuroscience and computer science efforts to model or engineer adaptive neural networks (be they biological or artificial) share the hard problem of optimally individualized adjustment of very large numbers of what both fields know as "synaptic weights". At the heart of this challenge is "credit assignment", that is, the assignment of "credit" (or "blame") to guide the strengthening (or weakening) of the small subset of synapses that actually contribute differentially to success (or failure) in a given perceptual, mnemonic or motor task. Neuroscientists struggle with the credit assignment problem as they search for biological learning rules. Computer scientists are driven by a quest for greater computational efficiency in training artificial networks and the prospect that evolution may have developed superior strategies. One concept that has come into prominence as a candidate biologically plausible solution to the credit assignment problem is that of modulated "Hebbian" or "spike-timing-dependent" plasticity (STDP) (Bengio et al., 2016; Dan and Poo, 2006; Farries and Fairhall, 2007; Florian, 2007; Frémaux and Gerstner, 2016; Froemke, 2015; Izhikevich, 2007; Marblestone et al., 2016; Pawlak et al., 2010; Poo et al., 2016; Roelfsema and Holtmaat, 2018; Xie and Seung, 2003) While most biological studies of modulated STDP so far have focused on the monoamine neuromodulator dopamine (Brzosko et al., 2019; Izhikevich, 2007; Kuśmierz et al., 2017; Schultz, 2015) known commonalities of signal transduction downstream from widely varying GPCRs suggest that NP-GPCRs could play roles in credit assignment analogous to those postulated for dopamine-selective GPCRs (Hamilton et al., 2013; Roelfsema and Holtmaat, 2018) but see (Edelmann and Lessmann, 2011).

Deeper understanding of neuromodulation roles in adaptive cortical function seems certain to require a framework for integrating consideration of the panoply of possible activity-dependent modulatory networks with modulated excitatory and inhibitory synaptic networks. **Figure 8** conceptualizes one such framework schematically, using a common neurotaxonomy to integrate statistics of multiple neuromodulatory and multiple synaptic signaling networks. Panels **A-C** idealize a logic for prediction of neuropeptidergic connectivity statistics from transcriptomic data. Panel **D** cartoons the use of a common neurotaxonomy to integrate probabilistic NP network graphs (panels **B,C;** three in this case) and multiple synaptic networks (panels **E,F;** two in this case) into a single graph representing superimposed modulatory and synaptic network. The present analysis suggests that a more realistic materialization of the **Figure 8** schematic would involve approximately 100 neuron types and dozens of NPP and NP-GPCR genes. It would also require information that is presently unavailable about excitatory and inhibitory synaptic connectivity statistics in such a neurotaxonomic framework. It is very encouraging, however, that vigorous ongoing efforts (e.g., see (Daigle et al., 2018; Jonas and Kording, 2015; Swanson and Lichtman, 2016; Tasic, 2018; Zeng and Sanes, 2017) suggest that such information is on the way. A view of cortical circuitry as a superimposition of multiple modulatory and synaptic networks, linked by a common neurotaxonomy as idealized in **Figure 8**, may prove essential to fathoming the interplay of slow neuromodulation and fast synaptic signaling necessary for adaptive cortical function.

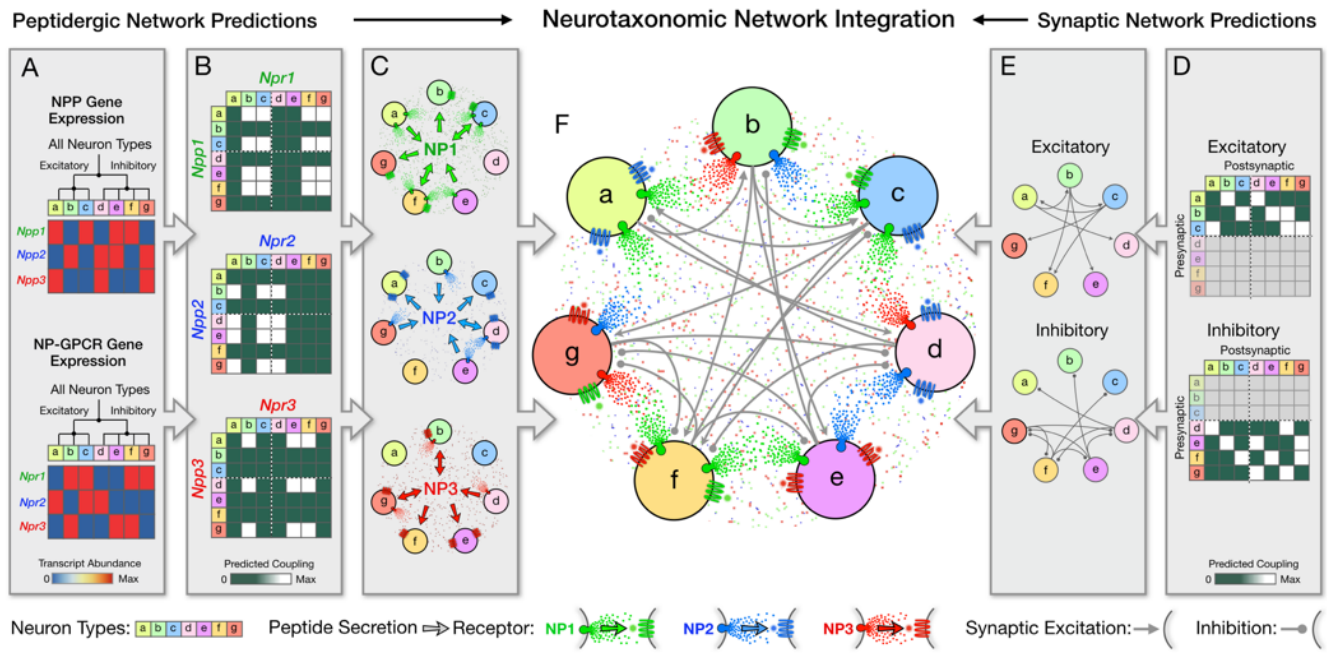


Figure 8. Neurotaxonomy offers a framework for integrating statistical descriptions of multiple modulatory and synaptic networks, as schematized here for purely fictitious neuron types, transcriptomic and connectomic data. Multiple directed network graphs are predicted here for peptidergic networks from transcriptomic data (A-C) and for synaptic networks from connectomic data (D,E) to predict a modulated synaptic network (F). (A) Transcriptional heat maps representing expression of three fictitious NPP genes (Npp1,2,3) and three cognate NP-GPCR genes (Npr1,2,3) across a neurotaxonomy comprising seven fictitious neuron types (a-c excitatory; d-g inhibitory). (B) Adjacency matrices derived from expression data in (A) as outer products of column and row factor vectors representing NPP and NP-GPCR expression, respectively. (C) Directed network graphs representing the same three NP networks, diagramming paracrine coupling by three neuropeptides (NP1,2,3) with routing of broadcast diffusive signals determined by differential expression of peptide-receptor pairings. (D) Neurotaxonomic adjacency matrices expressing excitatory and inhibitory synaptic connection statistics. (E) Synaptic network graphs derived from (D). (F) Directed multigraph illustrating use neurotaxonomy to integrate the three modulatory graphs and two synaptic connectivity graphs.

Implications for psychopharmacology

Neuropeptidergic signaling molecules have long beguiled as potential neuropsychiatric drug targets (Hökfelt et al., 2003; Hoyer and Bartfai, 2012). There seems to be some disappointment, however, in the returns on what has been rumoured to be very large research investments. The present study raises the possibility that both NP-targeted drug discovery and the reproducibility of physiological experimentation have been hindered by the same uncharted multiplicity, cell-type-specificity and redundancy of NPP and NP-GPCR expression. By charting these waters, single-neuron transcriptomic analysis may improve the odds substantially for both reproducible research and psychiatric drug development.

Today's psychiatric pharmaceuticals almost all target signaling by the monoamine neuromodulators dopamine, serotonin, noradrenaline and/or histamine and their selective GPCR receptors (Data-Franco et al., 2017; Hamon and Blier, 2013; Millan et al., 2015; Urs et al., 2014). Because they are so numerous, neuropeptide signaling systems may be much more neuron-type specific than monoamines. Greater neuron-type-specificity may translate to NP-targeting drugs being less troubled by side-effects and compensation (Hoyer and Bartfai, 2012). Moreover, while GPCRs have long been known as among the most "druggable" of targets (Gurrath, 2001; Lundstrom, 2009), the "druggability" of GPCRs is currently advancing very rapidly due to advances in GPCR structural biology and molecular dynamic simulations (Hilger et al., 2018; Koehl et al., 2018; Weis and Kobilka, 2018). It seems likely that new knowledge of the neuron-type-specificity of NP signaling gene expression will substantially advance the development of NP-targeting pharmaceuticals.

Conclusions

Because single-cell RNA-seq data enable prediction of complete protein parts lists of individual neurons, they open powerful new perspectives on neuronal differentiation, function and network architectures. The power of these new perspectives has been further enhanced by parallel development of transcriptomic neurotaxonomies. Here we have exploited both a pioneering large-scale RNA-seq dataset and its data-driven neurotaxonomy to pursue a new perspective on local neuropeptidergic modulatory signaling in mouse cortex. This work has revealed a surprisingly highly structured and abundant expression of cortical NPP and NP-GPCR genes: dozens of neuropeptide signaling genes are expressed at very high levels in very distinctive and highly conserved patterns. While entirely consistent with previous bulk transcriptomic and proteomic observations, it is only with the advent of the RNA-seq combination of single-cell resolution with genomic depth that this extreme structure and abundance has come into focus. We have endeavored here to shape these findings into specific and testable peptidergic signaling predictions in the hopes of guiding fruitful experimentation based on emerging transcriptomic neurotaxonomies, new means for genetic access to specific neuron types and powerful new tools for biophysical analysis of neuropeptide actions. The observations presented here suggest the intriguing possibility that the homeostasis, modulation and plasticity of cortical circuitry may involve local neuropeptidergic signaling networks of previously unrecognized abundance and density.

Materials and Methods

Data and software resources

The present study is based on analysis of a resource single-cell mRNA-seq dataset acquired at the Allen Institute (Tasic et al., 2018) and available for download at *http://celltypes.brain-map.org/rnaseq/.* These RNA-seq data were acquired from a total of 22,439 isolated neurons, with detection of transcripts from a median of 9,462 genes per cell (min = 1,445; max = 15,338) and an overall total of 21,931 protein-coding genes detected. 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., harvested tissue specimens from a variety of transgenic mice expressing fluorescent proteins to enable enrichment of samples for neurons and for relatively rare neuron types by FACS sorting after dissociation. This enrichment procedure resulted, by design, in a disproportionate representation of GABAergic neurons, canonically ~20% of neurons (Sahara et al., 2012), such that the sampled neuron population is roughly half GABAergic (47%) and half glutamatergic (53%). The resource publication (Tasic et al., 2018) should be consulted for full details of neuronal sample and library preparation, sequencing and data processing. Source data, spreadsheets, R scripts and other code used to generate all tables and figures presented here are available at *https://github.com/AllenInstitute/PeptidergicNetworks*. A derived data set *np_gpcr_cpm.csv* was used for analyses summarizing CPM, region and metadata for the NPP and NP-GPCR genes. The primary Tasic 2018 data tables are available for download at *http://celltypes.brain-map.org/rnaseq/.*

Data metrics

The Tasic 2018 single-cell RNA-seq data tables report the abundance of transcripts from individual neurons in both "counts per million reads" (CPM) and "fragments per kilobase of exon per million reads mapped" (FPKM) units. Our analysis of this data compares gene expression levels quantitatively, with two distinct use cases: (1) comparisons across large sets of different genes, and (2) comparisons of the same gene across different individual cells, cell types and brain areas. We have relied upon FPKM data (Mortazavi et al., 2008; Pimentel, 2014), for use case 1 (i.e., the Table 1 and 2 comparisons across genes). For use case 2 (as in all figures below), we have preferred the CPM units, because these units were used to generate the Tasic 2018 neurotaxonomy. While choices between CPM and FPKM units here should have little impact upon outcomes, it would seem inconsistent to use FPKM units to compare across cell types discerned on the basis of CPM units.

The NP signaling genes upon which the present analysis focuses are 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 FPKM" (pFPKM) metric. This metric is generated by ranking single-cell FPKM values for a given gene across the entire population of 22,439 neurons sampled, then designating the FPKM value at the ascending 99.9th percentile point as pFPKM. This metric was designed to minimize effects of sporadic outliers and sample size while still closely approximating the actual peak expression value in even very small subsets of neurons expressing the gene in question. Figures 1A and 2A, and their Source data files provide very detailed additional information about the single-cell RNA-seq value distributions sampled by the pFPKM metrics.

Selection of the 18 NPP gene set

As noted in **Introduction**, usage and definitions of the term "neuropeptide" vary widely across the current literature. It therefore seems unwise at present to claim that any attempted consensus list would accurately circumscribe all neuropeptides. For the purposes of the present work, we have relied therefore on the reasonably exhaustive list of 96 classical and candidate human and mouse NPP genes put forth in a widely cited publication (Burbach, 2010) and related website (http://neuropeptides.nl/, last accessed 10 October 2019). To reconcile this list to current mouse gene nomenclature, we used both the HGNC nomenclature ((https://www.genenames.org/, last accessed 10 October 2019) and the Mouse Genome Database (MGD) (http://www.informatics.jax.org, last accessed 10 October 2019). The result is the list of 94 putative mouse NPP genes presented in **Supplementary file 3**, which also tabulates the pFPKM values

and percentile scores compiled for each NPP genes from the Tasic 2018 dataset. These 94 NPP genes were further segregated using a preliminary (early 2018) version of the Tasic 2018 neurotaxonomy to select NPP genes exhibiting median CPM expression levels >10 in one or more neuron type in VISp and ALM cortex. This screening resulted in the list of 39 such NPP genes represented in **Supplementary file 4**, with most exceeding the 10 CPM threshold by a large margin (observed range was 24-4100 CPM). **Supplementary file 4** also tabulates criteria that drove inclusion of only the 18 NPP genes represented in Table 1 while 21 other cortically expressed NPP genes were excluded. The 18 select NPP genes include all but two (*Edn3* and *Gal*) genes for which transcripts ranked in the top quintile by pFPKM of the 94 putative NPP genes as tabulated in **Supplementary file 3**.

Selection of the 29 NP-GPCR gene set

The 18 select NPP genes listed in **Table 1** were used to search manually for cognate NP-GPCRs expressed in mouse cortex, relying primarily on ligand/receptor pairing data retrieved from the IUPHAR/BPS Pharmacology website (http://www.guidetopharmacology.org, accessed in March, 2018) and the Tasic 2018 NP-GPCR expression data tabulated in **Supplementary file 5**. This process resulted in selection of the 29 mouse NP-GPCR genes listed in **Table 2**, which also lists for each the corresponding cognate NPP gene or genes used to root the search. The matching of NP-GPCR and NPP genes in **Table 2** neglects a few receptor/ligand pairings rated on the IUPHAR/BPS website as very low in affinity compared to primary pairings.

Autoencoder-based classifier development and evaluation methods:

Gene sets

Table of different sets of genes used for experiments shown in Figure 4:

Gene set	Description
NP47	The combined set of 18 NPPs and 29 NP-GPCRs
HE	6,083 genes selected based on maximum value across all neurons in the dataset
DE	4,020 differentially expressed genes for Tasic 2018 neurotaxonomy
DE47	47 most variable genes selected from the set of DE genes
Rand47	Random subsets of 47 genes drawn from the set of HE genes
Rand47 ExpMatched	Random subsets of 47 genes such that the maximum expression value approximately matches that of the NP genes

Autoencoder network architecture

Autoencoders are multi-layer, feedforward neural network models that consist of encoder/decoder subnetworks. In its basic realization, the encoder subnetwork learns to compress the high dimensional input into a low dimensional representation, from which the decoder subnetwork estimates the original input. We constructed a network with two autoencoders, with 8 hidden layers each. The architecture of the first autoencoder (HE Genes autoencoder, **Figures 4C**) is Input(6083) → Dropout(0.8) → Dense(100) → Dense(100) → Dense(100) → Dense(100) → Dense(d) → Batch Normalization (latent representation z_1) → Dense(100) → Dense(100) → Dense(100) → Dense(100) → Dense(6083), and the architecture of the second autoencoder (NP Genes autoencoder, Figure 4B) is Input(47) → Dropout(0) → Dense(x) → Dense(x) → Dense(x) → Dense(x) → Dense(d) → Batch Normalization (latent representation z_2) → Dense(x)

→ Dense(x) → Dense(x) → Dense(x) → Dense(47). The numbers in parentheses of Dense denote the number of fully connected units in those layers. All Dense layer units use the rectified linear (ReLU) function as the nonlinear transformation except for those in the Dense(d) layers, which do not use a nonlinear transformation. For results using the NP genes autoencoder x=50; tests with x=25, led to qualitatively similar results (not shown) and did not change overall conclusions of the analyses. The Dropout layer (Srivastava et al., 2014) is used with dropout probability = 0.8 to regularize the HE Genes autoencoder and prevent over-fitting. The numbers of input/output units in each network match the number of input genes. The two dimensional representations (d=2) shown in **Figures 4A-B**, and the five dimensional (d=5) representations used in **Figure 4D** are the outputs of the Batch Normalization layer (Loffe and Szegedy, 2015) for the respective networks. We determined the optimal latent space dimensionality d=5 for the quantitative analysis by varying the latent space dimensionality of the HE Genes network between 2 and 20 dimensions and choosing the value that maximized the QDA analysis-based cell type classification accuracy for the HE genes (see **Figure 4 – Figure supplement 2**).

Autoencoder training

Both autoencoder networks were trained using the backpropagation algorithm with the Adam optimizer (Kingma and Ba, 2014) and a batch size of 956. The HE genes autoencoder was trained for 50,000 epochs using the mean squared error between the input and the output layers as the loss function. The NP genes autoencoder was trained for 10,000 epochs using L=R+λC as the loss function, where R denotes the mean squared reconstruction loss as in the HE genes network, C denotes the penalty for mismatch between the latent representations, and λ=100 is the weighting scalar between the two terms. After training the HE genes network and obtaining the latent representation z_1 for each cell, C calculates the mean squared error between the latent representation of the NP genes network z_2 and z_1, while simultaneously normalizing variance along the narrowest direction for z_2. The two additive loss terms, R and C, together minimize the reconstruction error while attempting to match the representation learned using only the HE gene set. The same procedure was used for all small gene subsets including NP and random gene sets. Python implementations of the networks using the Tensorflow and Keras libraries are included in the code repository.

Quantifying abilities of gene sets to classify cell types

The neurotaxonomy of Tasic et al., 2018 defines hierarchical relationships of neuronal cell types. For each gene set, we used Quadratic Discriminant Analysis (QDA) to train multiple classifiers on the latent space representations to predict labels at different levels of the cell type hierarchy. The different levels (nodes) in the hierarchy were characterized in Tasic 2018 with a resolution index measure. Here we re-normalized that resolution index measure to have a value of 0.0 for the class of neurons (root node), and 1.0 for the 115 VISp+ALM cell types (leaf nodes, inset in **Figure 4D**). All intermediate nodes in the hierarchical classification tree have a positive resolution index that is less than 1.0. We used this property of nodes in the hierarchical classification tree to assign a resolution index (RI) value to each cell. The procedure starts with a classifier that was trained using all the leaf node labels, i.e. all the 115 VISp+ALM cell type labels. Test cells that are classified correctly at this level are assigned RI=1.0, which corresponds to the resolution index measure of the leaf nodes. Test cells that are incorrectly classified at this level of detail are re-assigned labels by a classifier that was trained on successively merged labels along the hierarchical tree till they are correctly classified. These cells receive the resolution index value of the node for which they are assigned the correct label. This procedure was performed using 13 fold cross validation for all the different gene sets, and the results were pooled.

Peptidergic Coupling Matrices

For a given cortical area $A \in \{ALM, VISp\}$, we denote by $NPP_A(g,t)$ the mean CPM expression matrix having entries NPP gene g and cell type *t.* Similarly, $NPGPCR_A(h,t)$ has as entries the expression of NP-GPCR gene *h* in type *t.* The *coupling matrix* $C^A_{(g,h)}$ of the pair *(g,h)* in area *A* is then defined $C^A_{(g,h)}(t,s) = \log_{10}(NPP_A(g,t) \cdot NPGPCR_A(h,s))$ for the fixed pair (g,h) in (NPP, NP-GPCR) as *t,s* range over all cell types in *A.* Matrices $C^A_{(g,h)}$ are formally the (square matrix) outer product $NPP_A \otimes NPGPCR_A$ then

presented in log10 units. Pooled representations are computed by averaging values of coupling matrices $C^{A}_{(g,h)}$ over 12 major cell types prior to rendering.

Transduction mode predictions

Peptidergic coupling matrices are summed, log10 scaled and maximum normalized independently according to Gi/o, Gs and Gq/11 family membership, then displayed in red, green and blue, respectively. Pooled representations are computed by averaging type-level data over subclasses before similar rendering.

Acknowledgements

We wish to thank the Allen Institute for Brain Science founder, Paul G. Allen, for his vision, encouragement and support. 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

Abbott LF, Regehr WG. 2004. Synaptic computation. *Nature* **431**:796–803. doi:10.1038/nature03010

Alexander SPH, Christopoulos A, Davenport AP, Kelly E, Marrion N V., Peters JA, Faccenda E, Harding SD, Pawson AJ, Sharman JL, Southan C, Davies JA. 2017. THE CONCISE GUIDE TO PHARMACOLOGY 2017/18: G protein-coupled receptors. *Br J Pharmacol* **174**:S17–S129. doi:10.1111/bph.13878

Banghart MR, He XJ, Sabatini BL. 2018. A Caged Enkephalin Optimized for Simultaneously Probing Mu and Delta Opioid Receptors. *ACS Chem Neurosci* **9**:684–690. doi:10.1021/acschemneuro.7b00485

Banghart MR, Williams JT, Shah RC, Lavis LD, Sabatini BL. 2013. Caged Naloxone Reveals Opioid Signaling Deactivation Kinetics. *Mol Pharmacol* **84**:687–695. doi:10.1124/mol.113.088096

Baraban SC, Tallent MK. 2004. Interneuron Diversity series: Interneuronal neuropeptides - Endogenous regulators of neuronal excitability. *Trends Neurosci* **27**:135–142. doi:10.1016/j.tins.2004.01.008

Bargmann CI. 2012. Beyond the connectome: How neuromodulators shape neural circuits. *BioEssays* **34**:458–465. doi:10.1002/bies.201100185

Batista-Brito R, MacHold R, Klein C, Fishell G. 2008. Gene expression in cortical interneuron precursors is prescient of their mature function. *Cereb Cortex* **18**:2306–2317. doi:10.1093/cercor/bhm258

Bengio Y, Lee D-H, Bornschein J, Mesnard T, Lin Z. 2016. Towards Biologically Plausible Deep Learning. doi:10.1007/s13398-014-0173-7.2

Borbély É, Scheich B, Helyes Z. 2013. Neuropeptides in learning and memory. *Neuropeptides* **47**:439–450. doi:10.1016/j.npep.2013.10.012

Brzosko Z, Mierau SB, Paulsen O. 2019. Neuromodulation of Spike-Timing-Dependent Plasticity: Past, Present, and Future. *Neuron* **103**:563–581. doi:10.1016/j.neuron.2019.05.041

Bucher D, Marder E. 2013. XSnapShot: Neuromodulation. *Cell* **155**:482-482.e1. doi:10.1016/j.cell.2013.09.047

Burbach JPH. 2011. What are neuropeptides? *Methods Mol Biol* **789**:1–36. doi:10.1007/978-1-61779-310-3_1

Burbach JPH. 2010. Neuropeptides from concept to online database www.neuropeptides.nl. *Eur J Pharmacol* **626**:27–48. doi:10.1016/j.ejphar.2009.10.015

Cadwell CR, Sandberg R, Jiang X, Tolias AS. 2017. Q&A: Using Patch-seq to profile single cells. *BMC Biol* **15**:1–7. doi:10.1186/s12915-017-0396-0

Chen Y, Saulnier JL, Yellen G, Sabatini BL. 2014. A PKA activity sensor for quantitative analysis of endogenous GPCR signaling via 2-photon FRET-FLIM imaging. *Front Pharmacol* **5 APR**:1–12. doi:10.3389/fphar.2014.00056

Crawley JN. 1985. Comparative Distribution of Cholecystokinin and Other Neuropeptides: Why is This Peptide Different from All Other Peptides? *Ann N Y Acad Sci* **448**:1–8. doi:10.1111/j.1749-6632.1985.tb29900.x

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, Fong O, Nguyen TN, Garren E, Lenz GH, Chavarha M, Pendergraft J, Harrington J, Hirokawa KE, Harris JA, Nicovich PR, McGraw MJ, Ollerenshaw DR, Smith KA, Baker CA, Ting JT, Sunkin SM, Lecoq J, Lin MZ, Boyden ES, Murphy GJ, da Costa NM, Waters J, Li L, Tasic B, Zeng H. 2018. A Suite of Transgenic Driver and Reporter Mouse Lines with Enhanced Brain-Cell-Type Targeting and Functionality. *Cell* **174**:465-480.e22. doi:10.1016/j.cell.2018.06.035

Dan Y, Poo M-M. 2006. Spike Timing-Dependent Plasticity: From Synapse to Perception. *Physiol Rev*

86:1033–1048. doi:10.1152/physrev.00030.2005

Data-Franco J, Singh A, Popovic D, Ashton M, Berk M, Vieta E, Figueira ML, Dean OM. 2017. Beyond the therapeutic shackles of the monoamines: New mechanisms in bipolar disorder biology. *Prog Neuro-Psychopharmacology Biol Psychiatry* **72**:73–86. doi:10.1016/j.pnpbp.2016.09.004

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

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 SL, Hof PR, Huang ZJ, Jones EG, Kawaguchi Y, Kisvárday Z, Kubota Y, Lewis D a, Marín O, Markram H, McBain CJ, Meyer HS, Monyer H, Nelson SB, Rockland K, Rossier J, Rubenstein JLR, Rudy B, Scanziani M, Shepherd GM, Sherwood CC, Staiger JF, Tamás G, Thomson A, Wang Y, Yuste R, Ascoli G a. 2013. New insights into the classification and nomenclature of cortical GABAergic interneurons. *Nat Rev Neurosci* **14**:202–16. doi:10.1038/nrn3444

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, Jacob AL, Smith GB, Wilson DE, Kosche G, Kruglikov I, Rusielewicz T, Kotak VC, Mowery TM, Anderson SA, Callaway EM, Dasen JS, Fitzpatrick D, Fossati V, Long MA, Noggle S, Reynolds JH, Sanes DH, Rudy B, Feng G, Fishell G. 2016. A viral strategy for targeting and manipulating interneurons across vertebrate species. *Nat Neurosci* **19**:1743–1749. doi:10.1038/nn.4430

Ding K, Han Y, Seid TW, Buser C, Karigo T, Zhang S, Dickman DK, Anderson DJ. 2019. Imaging neuropeptide release at synapses with a genetically engineered reporter. *Elife* **8**:1–15. doi:10.7554/elife.46421

Edelmann E, Lessmann V. 2011. Dopamine modulates spike timing-dependent plasticity and action potential properties in CA1 pyramidal neurons of acute rat hippocampal slices. *Front Synaptic Neurosci* **3**:1–16. doi:10.3389/fnsyn.2011.00006

Elphick MR, Mirabeau O, Larhammar D. 2018. Evolution of neuropeptide signalling systems. *J Exp Biol* **221**:jeb151092. doi:10.1242/jeb.151092

Fan X, Dong J, Zhong S, Wei Y, Wu Q, Yan L, Yong J, Sun L, Wang Xiaoye, Zhao Y, Wang W, Yan J, Wang Xiaoqun, Qiao J, Tang F. 2018. Spatial transcriptomic survey of human embryonic cerebral cortex by single-cell RNA-seq analysis. *Cell Res* **28**:730–745. doi:10.1038/s41422-018-0053-3

Farries MA, Fairhall AL. 2007. Reinforcement Learning With Modulated Spike Timing Dependent Synaptic Plasticity. *J Neurophysiol* **98**:3648–3665. doi:10.1152/jn.00364.2007

Férézou I, Hill EL, Cauli B, Gibelin N, Kaneko T, Rossier J, Lambolez B. 2007. Extensive overlap of mu-opioid and nicotinic sensitivity in cortical interneurons. *Cereb Cortex* **17**:1948–1957. doi:10.1093/cercor/bhl104

Fishell G, Kepecs A. 2019. Interneuron Types as Attractors and Controllers 1–50.

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

Földy C, Darmanis S, Aoto J, Malenka RC, Quake SR, Südhof TC. 2016. Single-cell RNAseq reveals cell adhesion molecule profiles in electrophysiologically defined neurons. *Proc Natl Acad Sci* **113**:E5222–E5231. doi:10.1073/pnas.1610155113

Frémaux N, Gerstner W. 2016. Neuromodulated Spike-Timing-Dependent Plasticity, and Theory of Three-Factor Learning Rules. *Front Neural Circuits* **9**. doi:10.3389/fncir.2015.00085

Froemke RC. 2015. Plasticity of Cortical Excitatory-Inhibitory Balance. *Annu Rev Neurosci* **38**:195–219. doi:10.1146/annurev-neuro-071714-034002

Fu A, Pachter L. 2016. Estimating intrinsic and extrinsic noise from single-cell gene expression

measurements. *Stat Appl Genet Mol Biol 2016* **15**:447–471. doi:10.1515/sagmb-2016-0002

Gallopin T, Geoffroy H, Rossier J, Lambolez B. 2006. Cortical sources of CRF, NKB, and CCK and their effects on pyramidal cells in the neocortex. *Cereb Cortex* **16**:1440–1452. doi:10.1093/cercor/bhj081

Gokce O, Stanley GM, Treutlein B, Neff NF, Camp JG, Malenka RC, Rothwell PE, Fuccillo M V., Südhof TC, Quake SR. 2016. Cellular Taxonomy of the Mouse Striatum as Revealed by Single-Cell RNA-Seq. *Cell Rep* **16**:1126–1137. doi:10.1016/j.celrep.2016.06.059

Gomtsian L, Bannister K, Eyde N, Robles D, Dickenson AH, Porreca F, Navratilova E. 2018. 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* **159**:1. doi:10.1097/j.pain.0000000000001355

Gonzalez-Suarez AD, Nitabach MN. 2018. Peptide-Mediated Neurotransmission Takes Center Stage. *Trends Neurosci* **41**:325–327. doi:10.1016/j.tins.2018.03.013

Grimmelikhuijzen CJP, Hauser F. 2012. Mini-review: The evolution of neuropeptide signaling. *Regul Pept* **177**:S6–S9. doi:10.1016/j.regpep.2012.05.001

Gurrath M. 2001. Peptide-Binding G Protein-Coupled Receptors: New Opportunities for Drug Design. *Curr Med Chem* **8**:1605–1648.

Hackley CR, Mazzoni EO, Blau J. 2018. cAMPr: A single-wavelength fluorescent sensor for cyclic AMP. *Sci Signal* **11**. doi:10.1126/scisignal.aah3738

Haider RS, Godbole A, Hoffmann C. 2019. To sense or not to sense—new insights from GPCR-based and arrestin-based biosensors. *Curr Opin Cell Biol* **57**:16–24. doi:10.1016/j.ceb.2018.10.005

Hamilton TJ, Xapelli S, Michaelson SD, Larkum ME, Colmers WF. 2013. Modulation of Distal Calcium Electrogenesis by Neuropeptide Y1 Receptors Inhibits Neocortical Long-Term Depression. *J Neurosci* **33**:11184–11193. doi:10.1523/jneurosci.5595-12.2013

Hamm HE. 1998. The Many Faces of G Protein Signaling. *J Biol Chem* **273**:669–672.

Hamon M, Blier P. 2013. Monoamine neurocircuitry in depression and strategies for new treatments. *Prog Neuro-Psychopharmacology Biol Psychiatry* **45**:54–63. doi:10.1016/j.pnpbp.2013.04.009

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, Mallar Chakravarty M, Chapin M, Chong J, Dalley RA, Daly BD, Dang C, Datta S, Dee N, Dolbeare TA, Faber V, Feng D, Fowler DR, Goldy J, Gregor BW, Haradon Z, Haynor DR, Hohmann JG, Horvath S, Howard RE, Jeromin A, Jochim JM, Kinnunen M, Lau C, Lazarz ET, Lee C, Lemon TA, Li L, Li Y, Morris JA, Overly CC, Parker PD, Parry SE, Reding M, Royall JJ, Schulkin J, Sequeira PA, Slaughterbeck CR, Smith SC, Sodt AJ, Sunkin SM, Swanson BE, Vawter MP, Williams D, Wohnoutka P, Ronald Zielke H, Geschwind DH, Hof PR, Smith SM, Koch C, Grant SG, Jones AR. 2012. An anatomically comprehensive atlas of the adult human brain transcriptome. *Nature* **489**:391–399. doi:10.1038/nature11405

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. 2016. Strategies and Tools for Combinatorial Targeting of GABAergic Neurons in Mouse Cerebral Cortex. *Neuron* **92**:555. doi:10.1016/j.neuron.2016.10.009

Hilger D, Masureel M, Kobilka BK. 2018. Structure and dynamics of GPCR signaling complexes. *Nat Struct Mol Biol* **25**:4–12. doi:10.1038/s41594-017-0011-7

Hill SJ, Watson SP. 2018. Fluorescence Approaches Unravel Spatial and Temporal Aspects of GPCR Organisation, Location, and Intracellular Signalling. *Trends Pharmacol Sci* **39**:91–92. doi:10.1016/j.tips.2017.12.001

Hökfelt T, Bartfai T, Bloom F. 2003. Neuropeptides: Opportunities for drug discovery. *Lancet Neurol* **2**:463–472. doi:10.1016/S1474-4422(03)00482-4

Hökfelt T, Ögren SO, Xu ZQD. 2013. Classical Neurotransmitters and Neuropeptides, Second Edi. ed, Handbook of Biologically Active Peptides. Elsevier Inc. doi:10.1016/B978-0-12-385095-9.00251-7

Hoyer D, Bartfai T. 2012. Neuropeptides and neuropeptide receptors: Drug targets, and peptide and non-peptide ligands: A tribute to prof. dieter seebach. *Chem Biodivers* **9**:2367–2387. doi:10.1002/cbdv.201200288

Huang ZJ, Paul A. 2019. The diversity of GABAergic neurons and neural communication elements. *Nat Rev Neurosci*. doi:10.1038/s41583-019-0195-4

Huh D, Sejnowski TJ. 2017. Gradient Descent for Spiking Neural Networks.

Izhikevich EM. 2007. Solving the distal reward problem through linkage of STDP and dopamine signaling. *Cereb Cortex* **17**:2443–2452.

Jekely G. 2013. Global view of the evolution and diversity of metazoan neuropeptide signaling. *Proc Natl Acad Sci* **110**:8702–8707. doi:10.1073/pnas.1221833110

Jia S, Vaughan JC, Zhuang X. 2014. Isotropic three-dimensional super-resolution imaging with a self-bending point spread function. *Nat Photonics* **8**:302–306. doi:10.1038/nphoton.2014.13

Jonas E, Kording K. 2015. Automatic discovery of cell types and microcircuitry from neural connectomics. *Elife* **2015**:1–21. doi:10.7554/eLife.04250.001

Katz PS, Lillvis JL. 2014. Reconciling the deep homology of neuromodulation with the evolution of behavior. *Curr Opin Neurobiol* **29**:39–47. doi:10.1016/j.conb.2014.05.002

Kim JK, Kolodziejczyk AA, Illicic T, Teichmann SA, Marioni JC. 2015. Characterizing noise structure in single-cell RNA-seq distinguishes genuine from technical stochastic allelic expression. *Nat Commun* **6**. doi:10.1038/ncomms9687

Koch C, Segev I. 1998. Methods in Neuronal Modeling: From Ions to Networks, Computational neuroscience. MIT Press. doi:10.1109/MCISE.1999.743629

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, Kobilka BK. 2018. Structure of the μ-opioid receptor–Gi protein complex. *Nature* **558**:547–552. doi:10.1038/s41586-018-0219-7

Krishnan A, Schioth HB. 2015. The role of G protein-coupled receptors in the early evolution of neurotransmission and the nervous system. *J Exp Biol* **218**:562–571. doi:10.1242/jeb.110312

Kuffler SW, Jan LY, Jan YN. 1979. A peptide as a possible transmitter in sympathetic ganglia of the frog. *PNAS* **76**:1501–1505.

Kuśmierz Ł, Isomura T, Toyoizumi T. 2017. Learning with three factors: modulating Hebbian plasticity with errors. *Curr Opin Neurobiol* **46**:170–177. doi:10.1016/j.conb.2017.08.020

Lein E, Borm LE, Linnarsson S. 2017. The promise of spatial transcriptomics for neuroscience in the era of molecular cell typing. *Science (80-)*. doi:10.1126/science.aan6827

Lein E, Hawrylycz MJ, Ao N, Ayres M, Bensinger A, Bernard A, Boe AF, Boguski MS, Brockway KS, Byrnes EJ, Chen Lin, Chen Li, Chen TM, Chin MC, Chong J, Crook BE, Czaplinska A, Dang CN, Datta S, Dee NR, Desaki AL, Desta T, Diep E, Dolbeare TA, Donelan MJ, Dong HW, Dougherty JG, Duncan BJ, Ebbert AJ, Eichele G, Estin LK, Faber C, Facer BA, Fields R, Fischer SR, Fliss TP, Frensley C, Gates SN, Glattfelder KJ, Halverson KR, Hart MR, Hohmann JG, Howell MP, Jeung DP, Johnson RA, Karr PT, Kawal R, Kidney JM, Knapik RH, Kuan CL, Lake JH, Laramee AR, Larsen KD, Lau C, Lemon TA, Liang AJ, Liu Y, Luong LT, Michaels J, Morgan JJ, Morgan RJ, Mortrud MT, Mosqueda NF, Ng LL, Ng R, Orta GJ, Overly CC, Pak TH, Parry SE, Pathak SD, Pearson OC, Puchalski RB, Riley ZL, Rockett HR, Rowland SA, Royall JJ, Ruiz MJ, Sarno NR, Schaffnit K, Shapovalova N V., Sivisay T, Slaughterbeck CR, Smith SC, Smith KA, Smith BI, Sodt AJ, Stewart NN, Stumpf KR, Sunkin SM, Sutram M, Tam A, Teemer CD, Thaller C, Thompson CL, Varnam LR, Visel A, Whitlock RM, Wohnoutka PE, Wolkey CK, Wong VY, Wood M, Yaylaoglu MB, Young RC,

Youngstrom BL, Yuan XF, Zhang B, Zwingman TA, Jones AR. 2007. Genome-wide atlas of gene expression in the adult mouse brain. *Nature* **445**:168–176. doi:10.1038/nature05453

Lillicrap TP, Cownden D, Tweed DB, Akerman CJ. 2016. Random synaptic feedback weights support error backpropagation for deep learning. *Nat Commun* **7**:1–10. doi:10.1038/ncomms13276

Liu JJ, Sharma K, Zangrandi L, Chen C, Humphrey SJ, Chiu YT, Spetea M, Liu-Chen LY, Schwarzer C, Mann M. 2018. In vivo brain GPCR signaling elucidated by phosphoproteomics. *Science (80-)* **360**. doi:10.1126/science.aao4927

Livingston KE, Mahoney JP, Manglik A, Sunahara RK, Traynor JR. 2018. Measuring ligand efficacy at the mu-opioid receptor using a conformational biosensor. *Elife* **7**:1–23. doi:10.7554/eLife.32499

Ludwig M, Leng G. 2006. Dendritic peptide release and peptide-dependent behaviours. *Nat Rev Neurosci*. doi:10.1038/nrn1845

Lundstrom K. 2009. An Overview on GPCRs and Drug Discovery: Structure-Based Drug Design and Structural Biology on GPCRs In: Leifert WR, editor. G Protein-Coupled Receptors in Drug Discovery. Totowa, NJ: Humana Press. pp. 51–66. doi:10.1007/978-1-60327-317-6_4

Luo L, Callaway EM, Svoboda K. 2018. Genetic Dissection of Neural Circuits: A Decade of Progress. *Neuron* **98**:256–281. doi:10.1016/j.neuron.2018.03.040

Ma L, Jongbloets BC, Melander JB, Xiong W-H, Qin M, Lameyer TJ, Harrison MF, Zemelman B V., Mao T, Zhong H. 2018. A Highly Sensitive A-Kinase Activity Reporter for Imaging Neuromodulatory Events in Awake Mice. *Neuron* **99**:665–679. doi:10.1016/j.neuron.2018.07.020

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, Benucci A, Nagy A, Miyawaki A, Helmchen F, Empson RMM, Knöpfel T, Boyden ES, Reid RC, Carandini M, Zeng H, van der Bourg A. 2015. Transgenic mice for intersectional targeting of neural sensors and effectors with high specificity and performance. *Neuron* **85**:942–958. doi:10.1016/j.neuron.2015.02.022

Mains RE, Eipper BA. 2006. Chapter 18: Peptides In: Siegel GJ, Albers RW, Brady S, Price DL, editors. Basic Neurochemistry. Burlington, MA: Elsevier Academic Press. pp. 317–332.

Marblestone A, Wayne G, Kording K. 2016. Towards an integration of deep learning and neuroscience **10**:1–41. doi:10.3389/fncom.2016.00094

Marder E. 2012. Neuromodulation of Neuronal Circuits: Back to the Future. *Neuron* **76**:1–11. doi:10.1016/j.neuron.2012.09.010

Marder E, Goeritz ML, Otopalik AG. 2015. Robust circuit rhythms in small circuits arise from variable circuit components and mechanisms. *Curr Opin Neurobiol* **31**:156–163. doi:10.1016/j.conb.2014.10.012

Markram H, Gerstner W, Sjöström PJ. 2013. Spike Timing-Dependent Plasticity: A Comprehensive Overview, Frontiers in Synaptic Neuroscience. doi:10.3389/978-2-88919-043-0

Maximiliano José N, Hashikawa Y, Rudy B. 2018. Diversity and connectivity of layer 5 somatostatin-expressing interneurons in the mouse barrel cortex. *J Neurosci* **38**:2415–17. doi:10.1523/JNEUROSCI.2415-17.2017

Mccormick DA, Nusbaum MP. 2014. Editorial overview: Neuromodulation: Tuning the properties of neurons, networks and behavior. *Curr Opin Neurobiol* **29**:4–7. doi:10.1016/j.conb.2014.10.010

Mena JD, Selleck RA, Baldo BA. 2013. 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* **33**:18540–18552. doi:10.1523/JNEUROSCI.3323-12.2013

Mezey VA, Glasgow E, Kusano K, Chin H, Iii WSY, Gainer H. 1999. Single Cell Reverse Transcription-

Polymerase Chain Reaction Analysis of Rat Supraoptic Magnocellular Neurons : Neuropeptide Phenotypes and High Voltage-. *Endocrinoloy* **140**:5391–5401.

Millan MJ, Goodwin GM, Meyer-Lindenberg A, Ove Ogren S. 2015. Learning from the past and looking to the future: Emerging perspectives for improving the treatment of psychiatric disorders. *Eur Neuropsychopharmacol* **25**:599–656. doi:10.1016/j.euroneuro.2015.01.016

Mitre M, Minder J, Morina EX, Chao M V., Froemke RC. 2018. Oxytocin modulation of neural circuits. *Curr Top Behav Neurosci* **35**:31–53. doi:10.1007/7854_2017_7

Moffitt JR, Hao J, Bambah-mukku D, Lu T, Dulac C, Zhuang X. 2016. High-performance multiplexed fluorescence in situ hybridization in culture and tissue with matrix imprinting and clearing. *Proc Natl Acad Sci* **113**:201617699. doi:10.1073/pnas.1617699113

Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. 2008. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods* **5**:621–628.

Nadim F, Bucher D. 2014. Neuromodulation of neurons and synapses. *Curr Opin Neurobiol* **29**:48–56. doi:10.1016/j.conb.2014.05.003

Nässel DR. 2009. Neuropeptide signaling near and far: How localized and timed is the action of neuropeptides in brain circuits? *Invertebr Neurosci* **9**:57–75. doi:10.1007/s10158-009-0090-1

Nusbaum MP, Blitz DM, Marder E. 2017. Functional consequences of neuropeptide and small-molecule co-transmission. *Nat Rev Neurosci* **18**:389–403. doi:10.1038/nrn.2017.56

Okaty BW, Sugino K, Nelson SB. 2011. Cell Type-Specific Transcriptomics in the Brain. *J Neurosci* **31**:6939–6943. doi:10.1523/JNEUROSCI.0626-11.2011

Owen SF, Tuncdemir SN, Bader PL, Tirko NN, Fishell G, Tsien RW. 2013. Oxytocin enhances hippocampal spike transmission by modulating fast-spiking interneurons. *Nature* **500**:458–462. doi:10.1038/nature12330

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, Tian L. 2018. Ultrafast neuronal imaging of dopamine dynamics with designed genetically encoded sensors. *Science (80-)* **360**. doi:10.1126/science.aat4422

Paul A, Crow M, Raudales R, He M, Gillis J, Huang ZJ. 2017. Transcriptional Architecture of Synaptic Communication Delineates GABAergic Neuron Identity. *Cell* **171**:522-539.e20. doi:10.1016/j.cell.2017.08.032

Pawlak V, Wickens JR, Kirkwood A, Kerr JND. 2010. Timing is not everything: Neuromodulation opens the STDP gate. *Front Synaptic Neurosci* **2**:1–14. doi:10.3389/fnsyn.2010.00146

Pimentel H. 2014. What the FPKM? A review of RNA-Seq expression units. https://haroldpimentel.wordpress.com/2014/05/08/what-the-fpkm-a-review-rna-seq-expression-units/

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, Long J, Buzsáki G, Stevens C. 2016. What is memory? The present state of the engram. *BMC Biol* **14**:1–18. doi:10.1186/s12915-016-0261-6

Ratnayake K, Kankanamge D, Senarath K, Siripurapu P, Weis N, Tennakoon M, Payton JL, Karunarathne A. 2017. Measurement of GPCR-G protein activity in living cells, 2nd ed, Methods in Cell Biology. Elsevier Inc. doi:10.1016/bs.mcb.2017.07.008

Roelfsema PR, Holtmaat A. 2018. Control of synaptic plasticity in deep cortical networks. *Nat Rev Neurosci* **19**:166–180. doi:10.1038/nrn.2018.6

Rossier J, Chapouthier G. 1982. Brain opiates. *Endeavour* **6**:168–176. doi:10.1016/0160-9327(82)90072-

Russo AF. 2017. Overview of Neuropeptides: Awakening the Senses? *Headache* **57**:37–46. doi:10.1111/head.13084

Sahara S, Yanagawa Y, O'Leary DDM, Stevens CF. 2012. The Fraction of Cortical GABAergic Neurons Is Constant from Near the Start of Cortical Neurogenesis to Adulthood. *J Neurosci* **32**:4755–4761. doi:10.1523/jneurosci.6412-11.2012

Schultz W. 2015. Neuronal Reward and Decision Signals: From Theories to Data. *Physiol Rev* **95**:853–951. doi:10.1152/physrev.00023.2014

Shah S, Lubeck E, Schwarzkopf M, He T-F, Greenbaum A, Sohn CH, Lignell A, Choi HMT, Gradinaru V, Pierce NA, Cai L. 2016. Single-molecule RNA detection at depth by hybridization chain reaction and tissue hydrogel embedding and clearing. *Development* **143**:2862–2867. doi:10.1242/dev.138560

Shai AS, Larkum ME. 2017. Comment on: Towards deep learning with segregated dendrites. *Elife* **6**:1–2. doi:10.7554/eLife.22901

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. 2016. Comprehensive Classification of Retinal Bipolar Neurons by Single-Cell Transcriptomics. *Cell* **166**:1308-1323.e30. doi:10.1016/j.cell.2016.07.054

Smith SJ. 2018. Q&A: Array tomography. *BMC Biol* **16**:1–18. doi:10.1186/s12915-018-0560-1

Song S, Miller KD, Abbott LF. 2000. Competitive Hebbian learning through spike-timing-dependent synaptic plasticity. *Nat Neurosci* **3**:919–926. doi:10.1038/78829

Spangler SM, Bruchas MR. 2017. Optogenetic approaches for dissecting neuromodulation and GPCR signaling in neural circuits. *Curr Opin Pharmacol* **32**:56–70. doi:10.1016/j.coph.2016.11.001

Stoeber M, Jullié D, Lobingier BT, Laeremans T, Steyaert J, Schiller PW, Manglik A, von Zastrow M. 2018. A Genetically Encoded Biosensor Reveals Location Bias of Opioid Drug Action. *Neuron* 963–976. doi:10.1016/j.neuron.2018.04.021

Sugino K, Clark E, Schulmann A, Shima Y, Wang L, Hunt DL, Hooks BM, Tränkner D, Chandrashekar J, Picard S, Lemire AL, Spruston N, Hantman AW, Nelson SB. 2019. Mapping the transcriptional diversity of genetically and anatomically defined cell populations in the mouse brain. *Elife* **8**:1–29. doi:10.7554/eLife.38619

Sun F, Zeng J, Jing M, Zhou J, Feng J, Owen SF, Luo Y, Li F, Wang H, Yamaguchi T, Yong Z, Gao Y, Peng W, Wang L, Zhang S, Du J, Lin D, Xu M, Kreitzer AC, Cui G, Li Y. 2018. A Genetically Encoded Fluorescent Sensor Enables Rapid and Specific Detection of Dopamine in Flies, Fish, and Mice. *Cell* **174**:481-496.e19. doi:10.1016/j.cell.2018.06.042

Suter DM, Molina N, Gatfield D, Schneider K, Schibler U, Naef F. 2011. Mammalian genes are transcribed with widely different bursting kinetics. *Science (80-)* **332**:472–474. doi:10.1126/science.1198817

Swanson LW, Lichtman JW. 2016. From Cajal to Connectome and Beyond. *Annu Rev Neurosci* **39**. doi:10.1146/annurev-neuro-071714-033954

Syrovatkina V, Alegre KO, Dey R, Huang XY. 2016. Regulation, Signaling, and Physiological Functions of G-Proteins. *J Mol Biol* **428**:3850–3868. doi:10.1016/j.jmb.2016.08.002

Tasic B. 2018. Single cell transcriptomics in neuroscience: cell classification and beyond. *Curr Opin Neurobiol* **50**:242–249. doi:10.1016/j.conb.2018.04.021

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, Bernard A, Madisen L, Sunkin SM, Hawrylycz M, Koch C, Zeng H. 2016. Adult mouse cortical cell taxonomy revealed by single cell transcriptomics. *Nat Neurosci* **19**:335–346. doi:10.1038/nn.4216

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, Lathia K, Rimorin C, Tieu M, Larsen R, Casper T, Barkan E, Kroll M, Parry S, Shapovalova N V, Hirchstein D, Pendergraft J, Kim TK, Szafer A, Dee N, Groblewski P, Wickersham I, Cetin A, Harris JA, Levi BP, Sunkin SM, Madisen L, Daigle TL, Looger L, Bernard A, Phillips J, Lein E, Hawrylycz M, Svoboda K, Jones AR, Koch C, Zeng H. 2018. Shared and distinct transcriptomic cell types across neocortical areas. *Nature* **563**:72–78. doi:10.1101/229542

Telley L, Govindan S, Prados J, Stevant I, Nef S, Dermitzakis E, Dayer A, Jabaudon D. 2016. Sequential transcriptional waves direct the differentiation of newborn neurons in the mouse neocortex. *Science (80-)* **351**:1443–1446. doi:10.1126/science.aad8361

Tremblay R, Lee S, Rudy B. 2016. GABAergic Interneurons in the Neocortex: From Cellular Properties to Circuits. *Neuron* **91**:260–292. doi:10.1016/j.neuron.2016.06.033

Urs NM, Nicholls PJ, Caron MG. 2014. Integrated approaches to understanding antipsychotic drug action at GPCRs. *Curr Opin Cell Biol* **27**:56–62. doi:10.1016/j.ceb.2013.11.002

van den Pol AN. 2012. Neuropeptide Transmission in Brain Circuits. *Neuron* **76**:98–115. doi:10.1016/j.neuron.2012.09.014

Varoqueaux F, Fasshauer D. 2017. Getting Nervous: An Evolutionary Overhaul for Communication. *Annu Rev Genet Vol 51* **51**:455–476. doi:10.1146/annurev-genet-120116-024648

Vila-Porcile É, Xu ZQD, Mailly P, Nagy F, Calas A, Hökfelt T, Landry M. 2009. Dendritic synthesis and release of the neuropeptide galanin: Morphological evidence from studies on rat locus coeruleus neurons. *J Comp Neurol* **516**:199–212. doi:10.1002/cne.22105

von Diezmann A, Shechtman Y, Moerner WE. 2017. Three-Dimensional Localization of Single Molecules for Super-Resolution Imaging and Single-Particle Tracking. *Chem Rev* acs.chemrev.6b00629. doi:10.1021/acs.chemrev.6b00629

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. 2018. Three-dimensional intact-tissue sequencing of single-cell transcriptional states. *Science (80-)* **361**. doi:10.1126/science.aat5691

Wang Y, Wang M, Yin S, Jang R, Wang J, Xue Z, Xu T. 2015. NeuroPep: A comprehensive resource of neuropeptides. *Database* **2015**:1–9. doi:10.1093/database/bav038

Weis WI, Kobilka BK. 2018. The Molecular Basis of G Protein-Coupled Receptor Activation. *Annu Rev Biochem* **87**:897–919. doi:10.1146/annurev-biochem-060614-033910

Williams JT, Zieglgänsberger W. 1981. Neurons in the frontal cortex of the rat carry multiple opiate receptors. *Brain Res* **226**:304–308. doi:10.1016/0006-8993(81)91103-3

Xie X, Seung HS. 2003. Equivalence of backpropagation and contrastive Hebbian learning in a layered network. *Neural Comput* **15**:441–454. doi:10.1162/089976603762552988

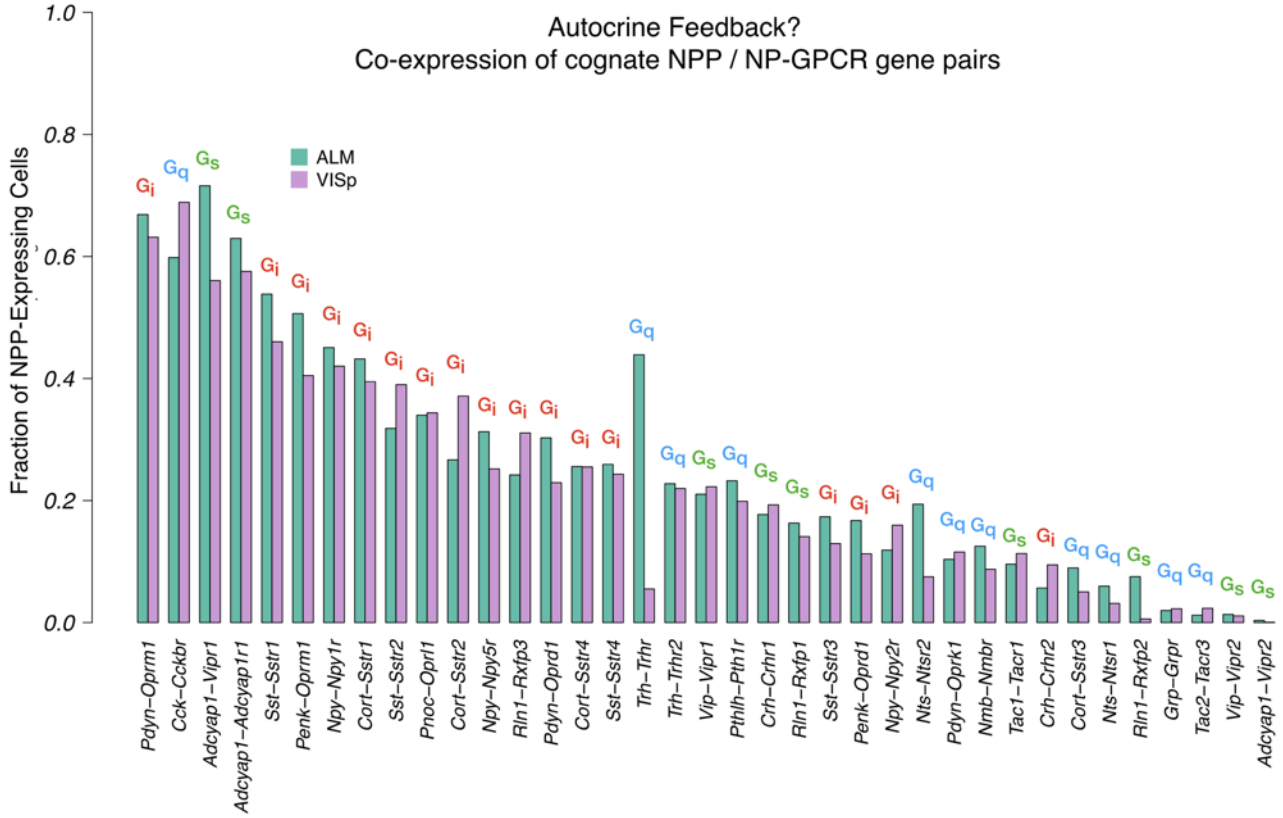
Xu X, Roby KD, Callaway EM. 2010. Immunochemical characterization of inhibitory mouse cortical neurons: Three chemically distinct classes of inhibitory cells. *J Comp Neurol* **518**:389–404. doi:10.1002/cne.22229

Zeisel A, Hochgerner H, Lönnerberg P, Johnsson A, Memic F, van der Zwan J, Häring M, Braun E, Borm LE, La Manno G, Codeluppi S, Furlan A, Lee K, Skene N, Harris KD, Hjerling-Leffler J, Arenas E, Ernfors P, Marklund U, Linnarsson S. 2018. Molecular Architecture of the Mouse Nervous System. *Cell* **174**:999-1014.e22. doi:10.1016/j.cell.2018.06.021

Zeng H, Sanes JR. 2017. Neuronal cell-type classification: challenges, opportunities and the path forward. *Nat Rev Neurosci*. doi:10.1038/nrn.2017.85

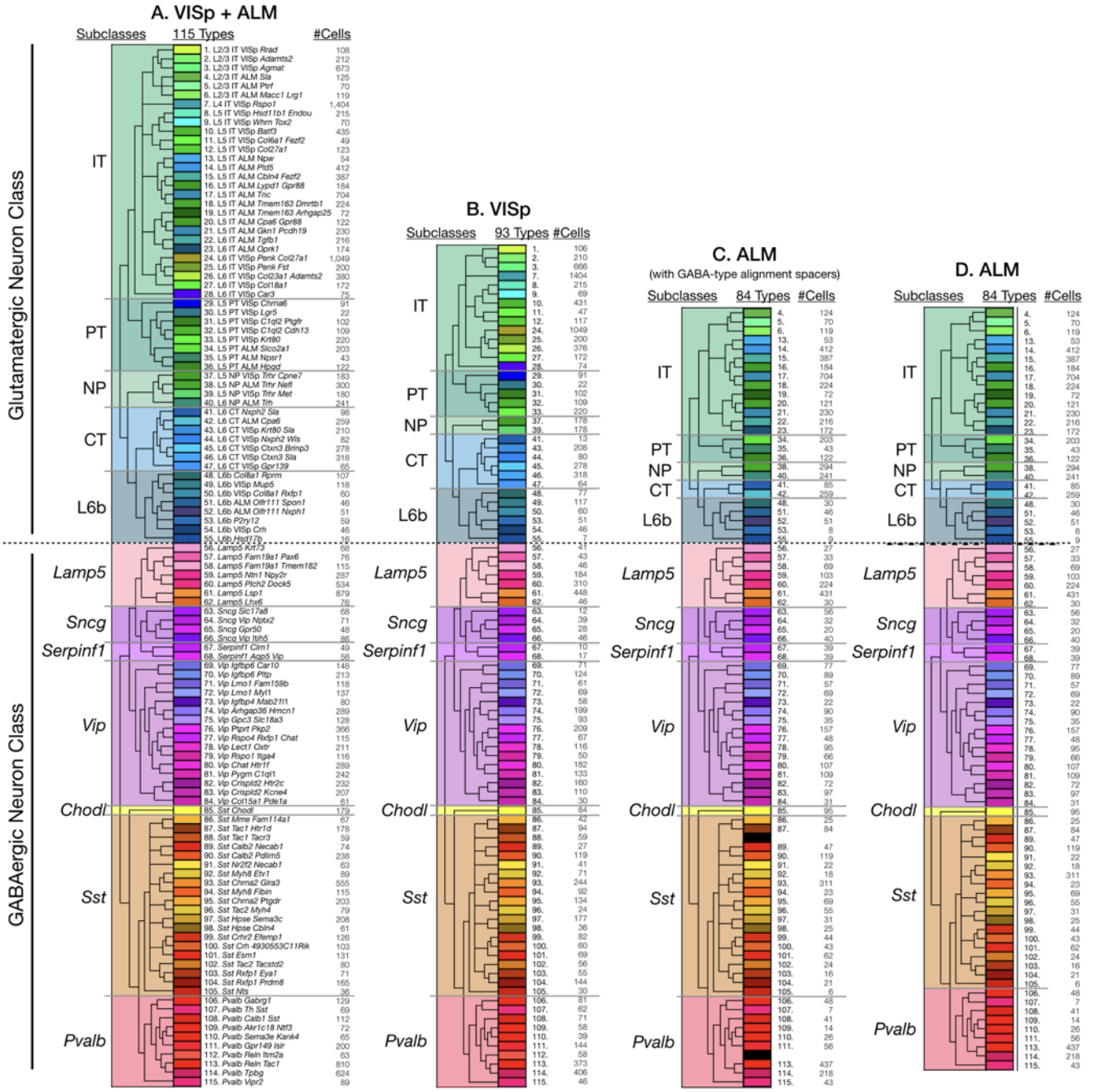
Figure Supplements

Figure 2 - figure supplement 1.



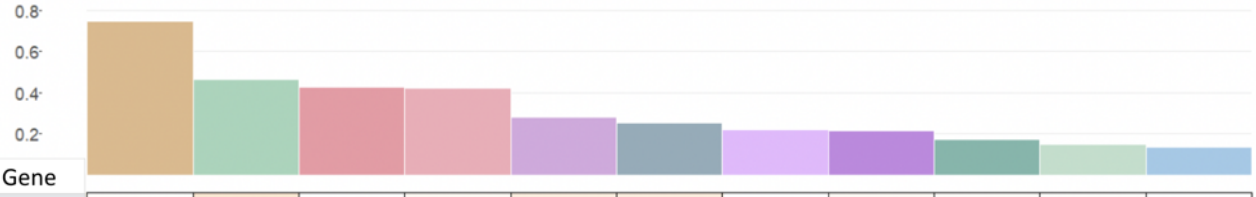
Co-expression of cognate NPP/NP-GPCR pairs. A large fraction of cortical neurons co-express cognate NPP/NP-GPCR pairs, and conservation of these fractions between cortical areas VISp and ALM argues in favor of some important underlying biology and possible relevance to function. Annotations of primary Gα family on this graph also hint that both positive and negative feedback mechanisms might be considered. These predictions of autocrine feedback, particularly the functionally negative feedback likely associated with Gi signaling, suggest the intriguing possibility that peptide signaling might underlie some forms of cell-autonomous neuronal homeostasis.

Figure 3 - Figure supplement 1 (referenced also from Figures 4-7)

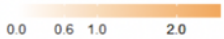


Tasic 2018 neurotaxonomy. Tasic and co-workers (Tasic et al., 2018, *Nature* **563**:72–78) generated the hierarchical taxonomy of glutamatergic and GABAergic neurons represented here by dimensionality reduction and iterative hierarchical clustering of data from single-cell RNA-Seq analysis of 22,439 mouse cortical neurons dissociated from areas VISp (13,491 neurons) and ALM (8,948 neurons). This neurotaxonomy is rendered here in four different forms as used in Figures 3-7 to specify Tasic 2018 neuronal cell "types" and/or "subclasses" as indicated by column headings. Each rendering represents a Tasic 2018 cladogram, 5 glutamatergic subclasses, 6 GABAergic subclasses (since Chodl is a singular type, it is not counted as a subclass) and some number of neuron types. The rectangle fill colors and orderings here are used consistently in Figures 3-7 to denote neuronal subclasses and types: the deep color fills for the small type rectangles are as introduced by the Tasic 2018 publication; the pastel color fills for the larger subclass rectangles are derived by averaging and fading the deeper color fills of the types within each subclass. The Tasic 2018 type numbers, names (panel A only) and counts of cells clustered to each type are tabulated here. (**A**) 11 subclasses and 115 types clustered from single-cell RNA-seq transcriptomes from cells aggregated across VISp and ALM areas. (**B**) 11 subclasses and 93 types found in VISp neurons. (**C**) 11 subclasses and 84 types found in ALM neurons. Because most GABAergic types (58 out of 60) are shared between VISp and ALM, spacers (black-filled rectangles) are placed in this rendering between types 87 and 89 and types 111 and 113 to maintain linear registration of GABAergic types between VISp and ALM types in spite of absence of types 88 and 112 in ALM. (**D**) 11 subclasses and 84 types found in ALM neurons (without registration-preserving spacers).

Figure 3 - Figure supplement 2

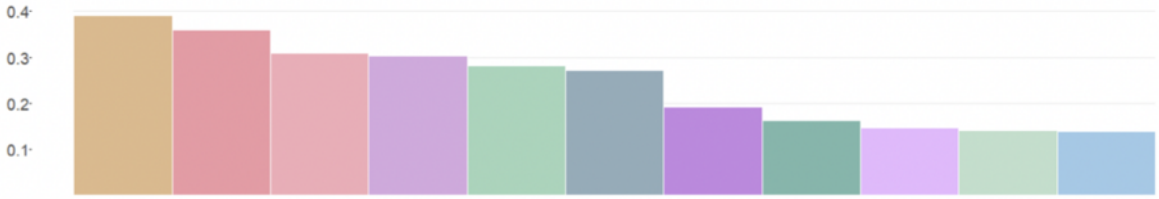


Gene	Sst	IT	Pvalb	Lamp5	Vip	L6b	Serpinf1	Sncg	PT	NP	CT
Nmb	0.06	0.64	0.02	0.16	0.45	0.46	0.02	0.14	0.1	0.07	0.02
Nts	2.03	0.46	1.07	0	0	0	0	0	0.24	0	0
Adcyap1	0	0.46	0	0.75	0	0.5	0	0	0.06	0.34	1.31
Rln1	0.13	0.05	0.42	0.39	0.11	0.37	0	0.1	0	0	0
Grp	0.95	0.71	0.94	0	0	0	0	0	0.42	0.51	0.45
Trh	0	0	0	0.62	0	0.58	0	0	0.02	1.73	0.26
Pnoc	0.59	0.01	0.42	0.1	0.28	0.23	0	0.05	0.06	0	0
Pthlh	0.8	0.68	0.27	0.89	0.25	0	2.01	0.01	0	0	0
Pdyn	0.79	1.28	0.01	0.01	0	0.43	0	0	0.58	0	0
Tac1	2.48	1.1	0.18	0.48	0.01	0.03	0.01	0.01	0.81	0	0.01
Cort	0.37	0.13	0.1	0.69	0.44	0.11	0.4	0.07	0.3	0	0.06
Crh	1.44	0.7	0.81	0.76	0.16	1.56	0.57	0.93	0.03	0.01	0.01
Penk	1.28	1.11	1.34	0.7	0.39	0.01	0	0.01	0.16	0	0.04
Cck	0.62	0.01	0.08	0.23	1.09	0.17	0	0.03	0.09	0	0.22
Tac2	1.61	0.3	0.01	1.54	0.54	0	0.58	1.04	0.02	0	0.01
Vip	0.01	0.01	0.02	0.04	0.05	0.01	0.26	1.2	0.09	0.01	0.01
Sst	0.03	0.08	1.46	0.01	0.01	0.02	0.02	0.05	0.1	0.01	0.03
Npy	0.21	0.67	0.57	0.24	1.31	0.04	0.1	0.21	0.03	0.02	0.02



NPP expression variation within subclasses. Coefficient of variation (c.v.=standard deviation/mean) of expression for each NPP gene for Tasic et al, 2018 cell types pooled to 12 major types, glutamatergic types (IT, PT, NP, CT, L6b) and GABAergic types (Lamp5, Sncg, Serpinf1, Vip, Chodl, Sst, Pvalb). Chodl does not appear in the figure here or in main text as there is only one Tasic cell type in this class. The type types are sorted from left to right in decreasing average c.v. The variability is a substantial fraction of the global variation across all types indicating complex combinatorial expression patterns at the level of the Tasic et al ontology and the need to consider this level of resolution. Outlier genes are colored more orange and examples labelled in Figure 3 of the main manuscript. Mean c.v is shown in the barplot on top.

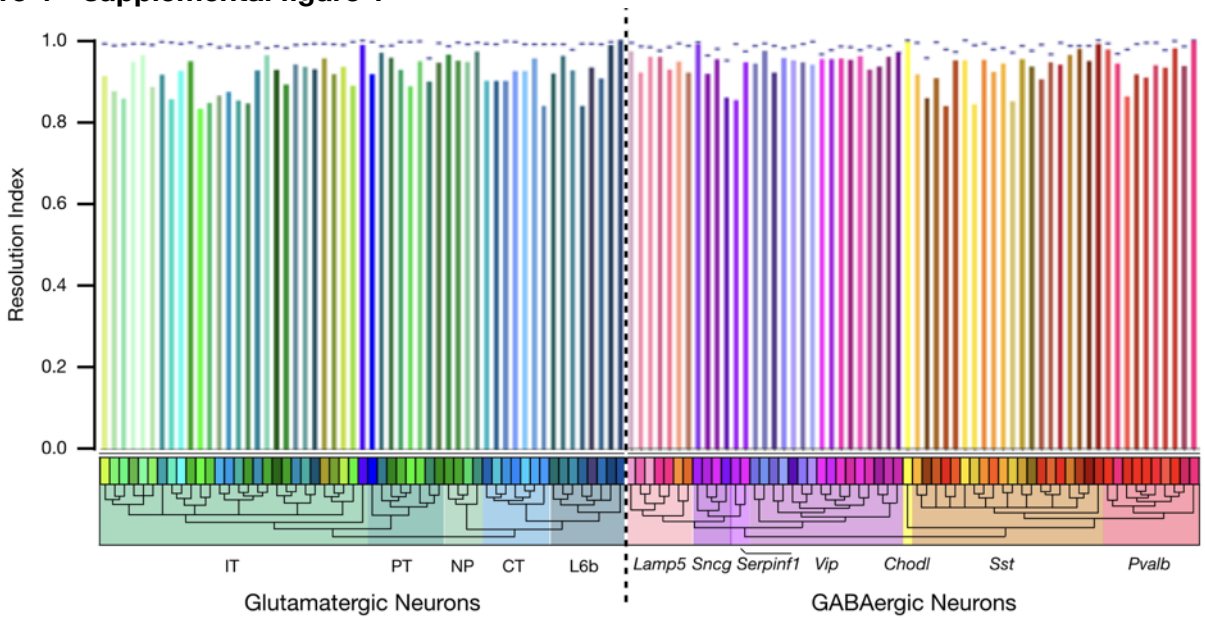
Figure 3 - figure supplement 3



Gene	*Sst*	*Pvalb*	*Lamp5*	*Vip*	IT	L6b	*Sncg*	PT	*Serpinf1*	NP	CT
Sstr3	0.16	0.01	0	0	0.02	0.06	0	0.32	0	0	0.16
Ntsr1	0	0	0.48	0	0.44	0.03	0.1	0.48	0	0	0.14
Oprd1	0.06	0.02	0	0	0.22	0	0	0.07	0	0	0.21
Rxfp2	0	1.15	0	0	0.88	0.88	0	0	0	0.44	0.25
Vipr2	0	1.34	0.02	0.63	0	0	0	0	0	0	0
Trhr2	0.22	0.06	0.03	0	0.31	0.44	0	0.37	0	0.53	0.09
Oprm1	0.06	1	0.2	0.49	0.22	0.02	0.2	0.07	0.18	0.32	0.06
Vipr1	0	0.02	0.02	0.5	0.02	0.2	0.57	0.06	0.38	0.06	0.03
Pth1r	0.21	0.17	0.46	0.44	0.39	0.34	0.14	0.1	0.45	0.02	0.29
Tacr1	0.34	0.79	1.42	0	0.15	0.27	0.04	0	0	0	0
Oprk1	0.54	0	0	0.15	1.71	0	0	0	0	0	0
Tacr3	1.38	1.45	0	0	0.4	0.22	0	0.01	0	0	0
Crhr2	0.89	0.19	0	0.71	0	0	0	0	0	0	0
Oprl1	0.01	0.03	0.03	0.89	0.03	0	0.24	0.01	1.22	0.01	0.02
Rxfp3	0.78	0	0.85	0.54	0	0	0.23	0	0	0	0
Crhr1	0.01	0	0	0.52	0.08	0.55	0	0.02	0	0	0.07
Adcyap1r1	0.03	0	0.01	0.02	0.05	0.01	0.01	1.03	0	0.01	0.71
Sstr1	0.16	0.97	0.13	1.13	0.23	0.96	0.07	0.49	0.26	0.24	0.06
Trhr	0	0	0.14	0	0	0.86	0.28	0.33	0	0.05	0.07
Sstr4	0.34	0.15	0.54	0.3	0.18	0.04	0.54	0.02	0	0.05	0.01
Rxfp1	1.23	0.6	1.17	0.7	1.13	0.29	0.75	0.55	1.04	0.82	0.62
Nmbr	0.84	0.63	0.59	0.39	0.19	0.55	0.1	0.31	0	0	0.68
Npy5r	0.51	0.5	0.36	0.02	0.08	0.44	0.02	0.16	0.05	0.25	0.16
Ntsr2	0.33	0.06	0.41	0.53	0.09	0.69	0.07	0.12	0.04	0.12	0.3
Cckbr	0.57	0.19	0.53	0	0.02	0	0.41	0.03	0.01	0.09	0
Grpr	0.58	0	0	0.03	0	0	0.43	0	0.45	0	0
Npy1r	0.86	0	0.34	0.02	0.02	0.44	0.01	0.16	0.06	0.43	0.04
Npy2r	0.72	0	1.14	0.58	0.9	0	0.92	0	0	0	0
Sstr2	0.52	1.12	0.09	0.17	0.43	0.6	0.42	0.04	0.09	0.63	0.02

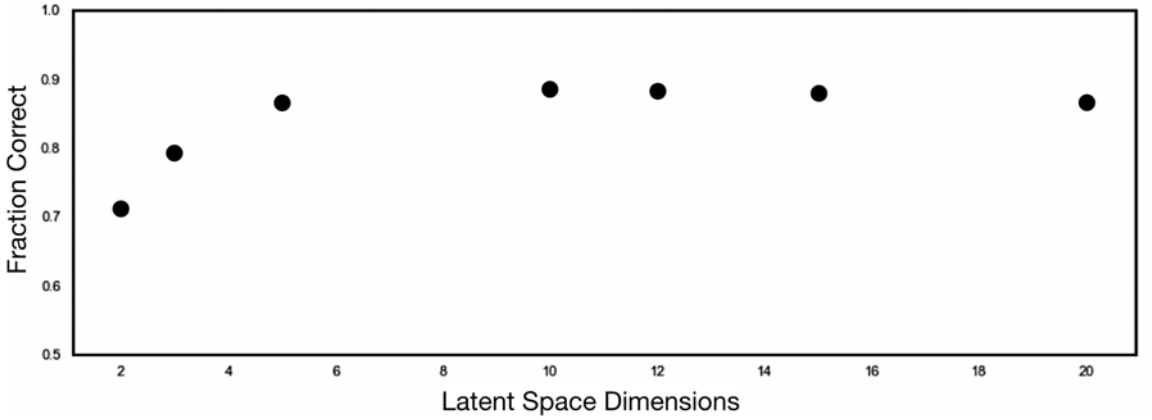
NP-GPCR expression variation within subclasses. Coefficient of variation (c.v.=standard deviation/mean) of expression for each NP-GPCR gene for Tasic et al, 2018 cell types pooled to 12 major types, glutamatergic types (IT, PT, NP, CT, L6b) and GABAergic types (Lamp5, Sncg, Serpinf1, Vip, Chodl, Sst, Pvalb). Mean c.v is a higher percentage of total c.v. for NP-GPCR compared to NPP genes.

Figure 4 – supplemental figure 1



Autoencoder expression clustering performance by neuron type. Results here compare classification based on 6,083 HE genes (dashed line) with results based on 47 NP genes (colored vertical bars). The resolution index (RI), as described in connection with Figure 4, generates a measure of autoencoder clustering agreement with the Tasic 2018 neurotaxonomy. Figure 4D represents resolution index distributions across all glutamatergic and GABAergic neurons represented in the Tasic 2018 dataset. This population is weighted, however, by the Tasic 2018 sampling process and does not represent the actual distribution of neuron types in cortex. As a check for possible anomalies related to this sampling process, we averaged resolution indices separately for each Tasic 2018 neuron type, as shown here. The single-type average resolution indices for both HE and NP gene sets are high across all types (>0.95 for HE and >0.8 for NP), confirming the adequacy of both gene sets in identifying all of the Tasic 2018 neuron types. Interestingly, this metric suggests that clustering performance using the 6,083 HE gene set is slightly worse for GABAergic than for glutamatergic types, while the reverse is true for the 47 NP gene set. Both effects are, however, quite small.

Figure 4 – supplemental figure 2



1``Determination of optimal autoencoder latent space dimensionality. 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.

Figure 6 – figure supplement 1A

Neurotaxonomic coupling predictions for 37 cognate NPP/NP-GPCR pairs. Predictions for all 37 cognate pairs in both VISp and ALM cortical areas are grouped and color-coded here and in the following figure supplements 1B-1H according to G_α signal transduction family. All predictions were generated as described in connection with Fig. 6 and represented as in Figs. 6C-E for neuron types and in 6F for subclasses. Numbers at upper right indicate the maximum value to which each plot was normalized, in units of CPM*CPM.

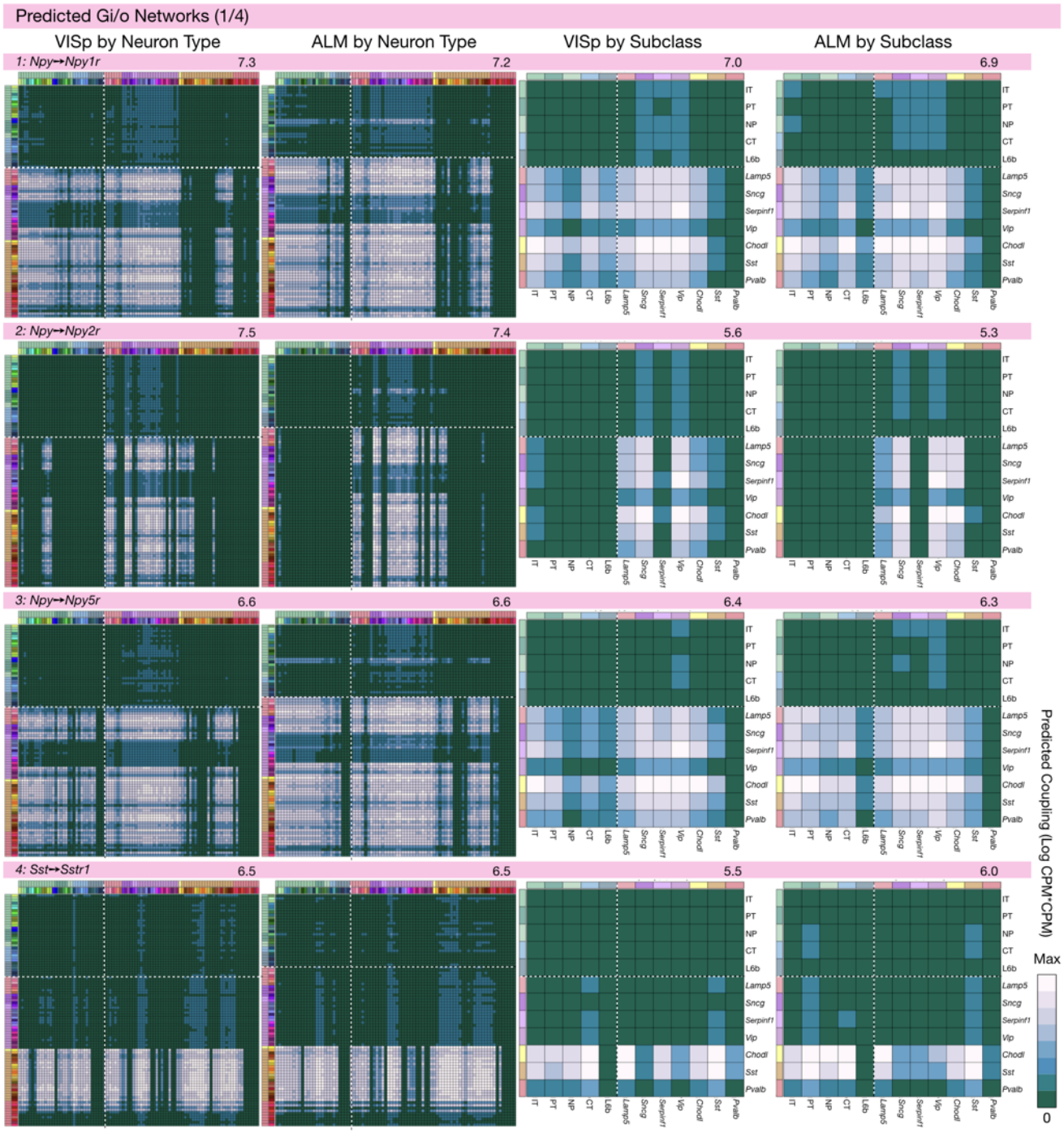


Figure 6 – figure supplement 1B (see legend, figure supplement 1A)

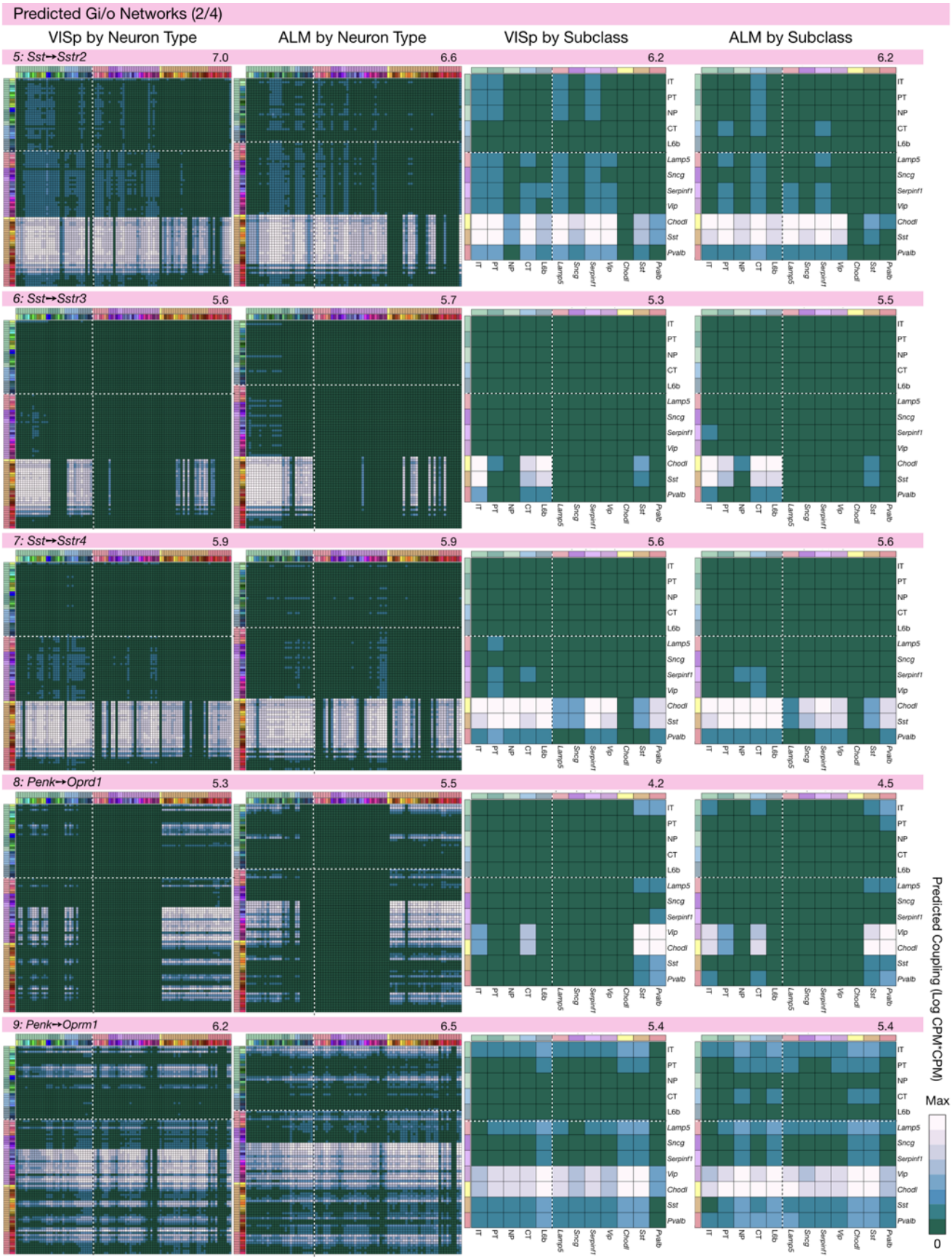


Figure 6 – figure supplement 1C (see legend, figure supplement 1A)

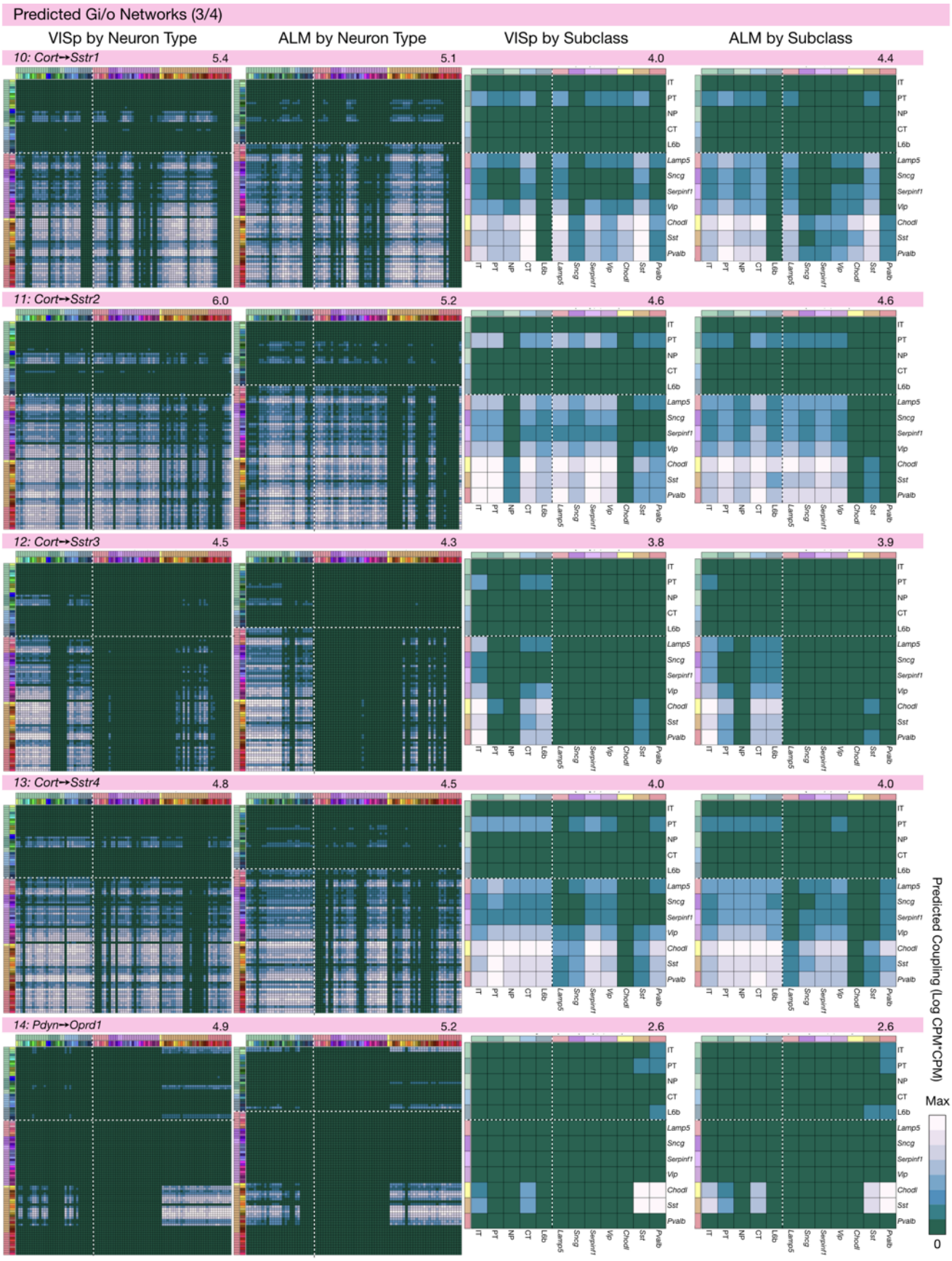


Figure 6 – figure supplement 1D (see legend, figure supplement 1A)

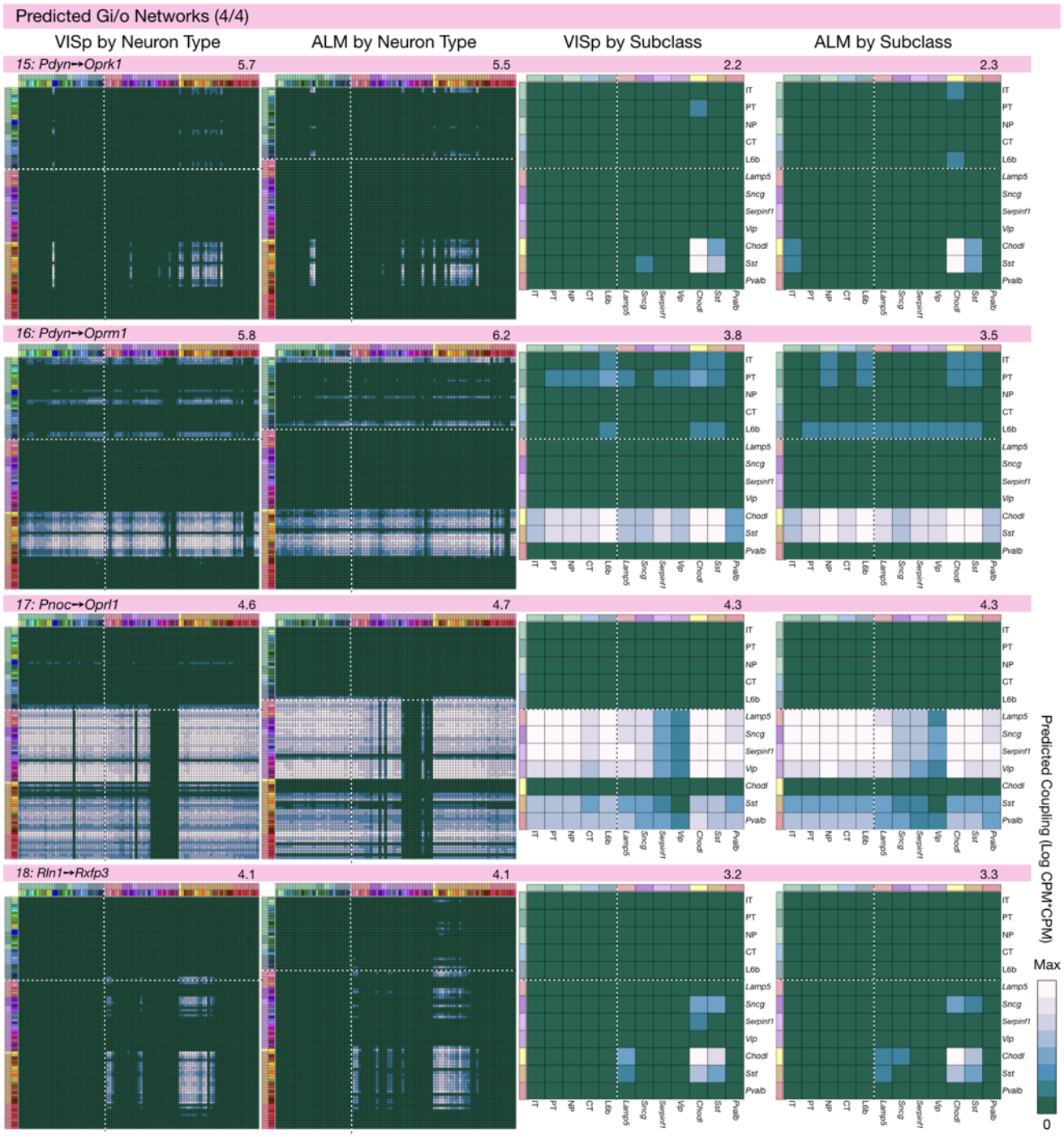


Figure 6 – figure supplement 1E (see legend, figure supplement 1A)

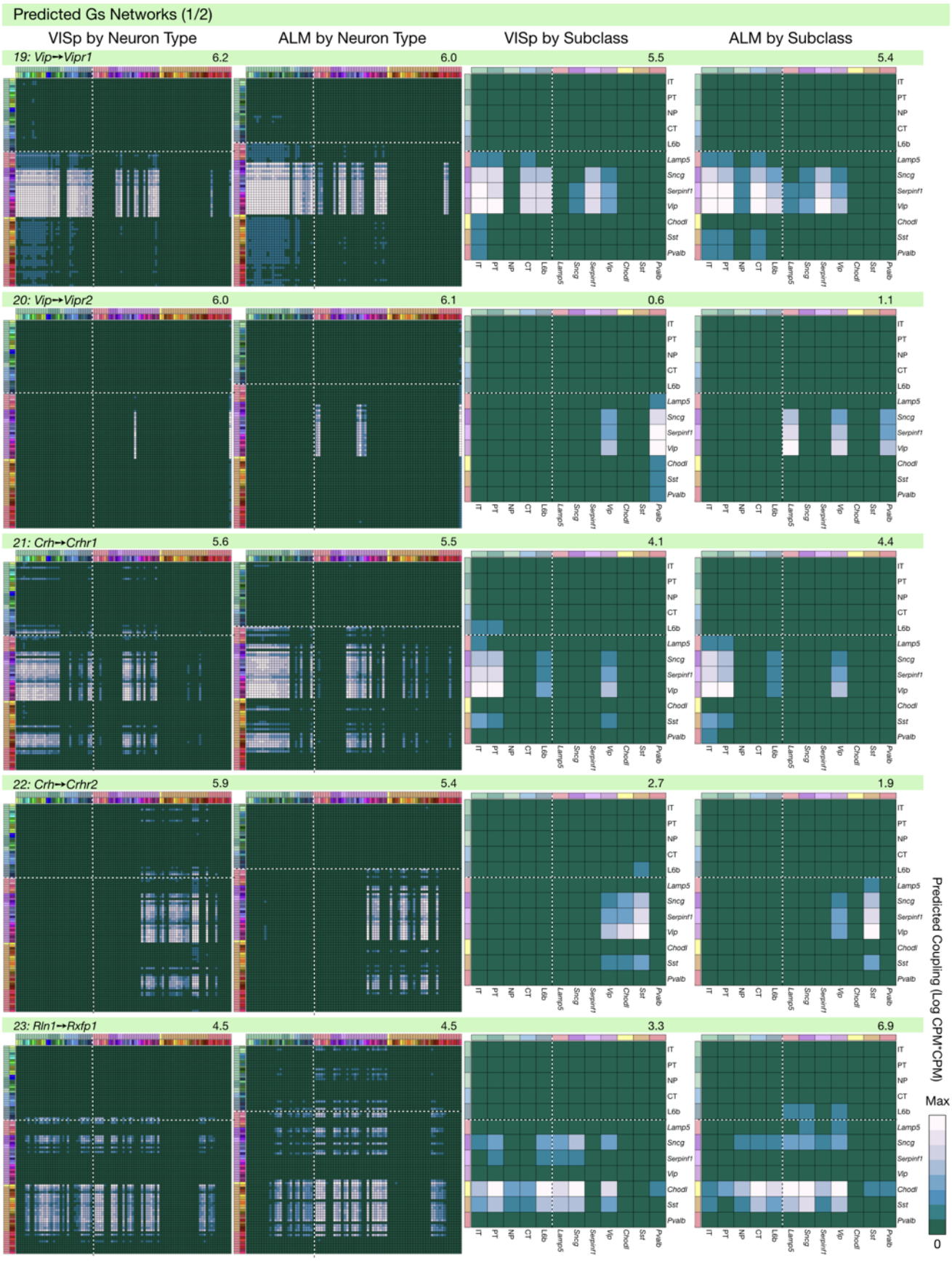


Figure 6 – figure supplement 1F (see legend, figure supplement 1A)

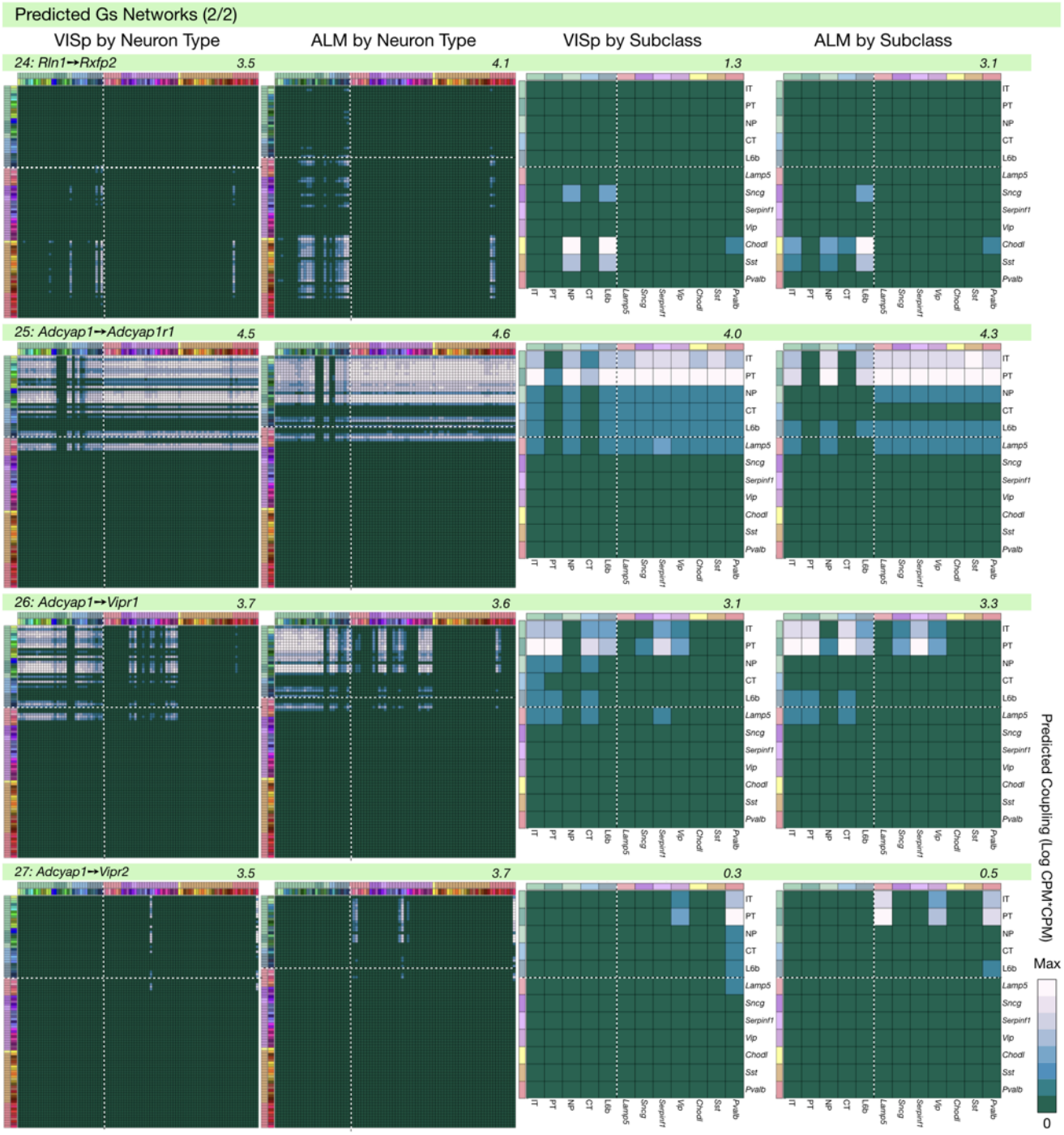


Figure 6 – figure supplement 1G (see legend, figure supplement 1A)

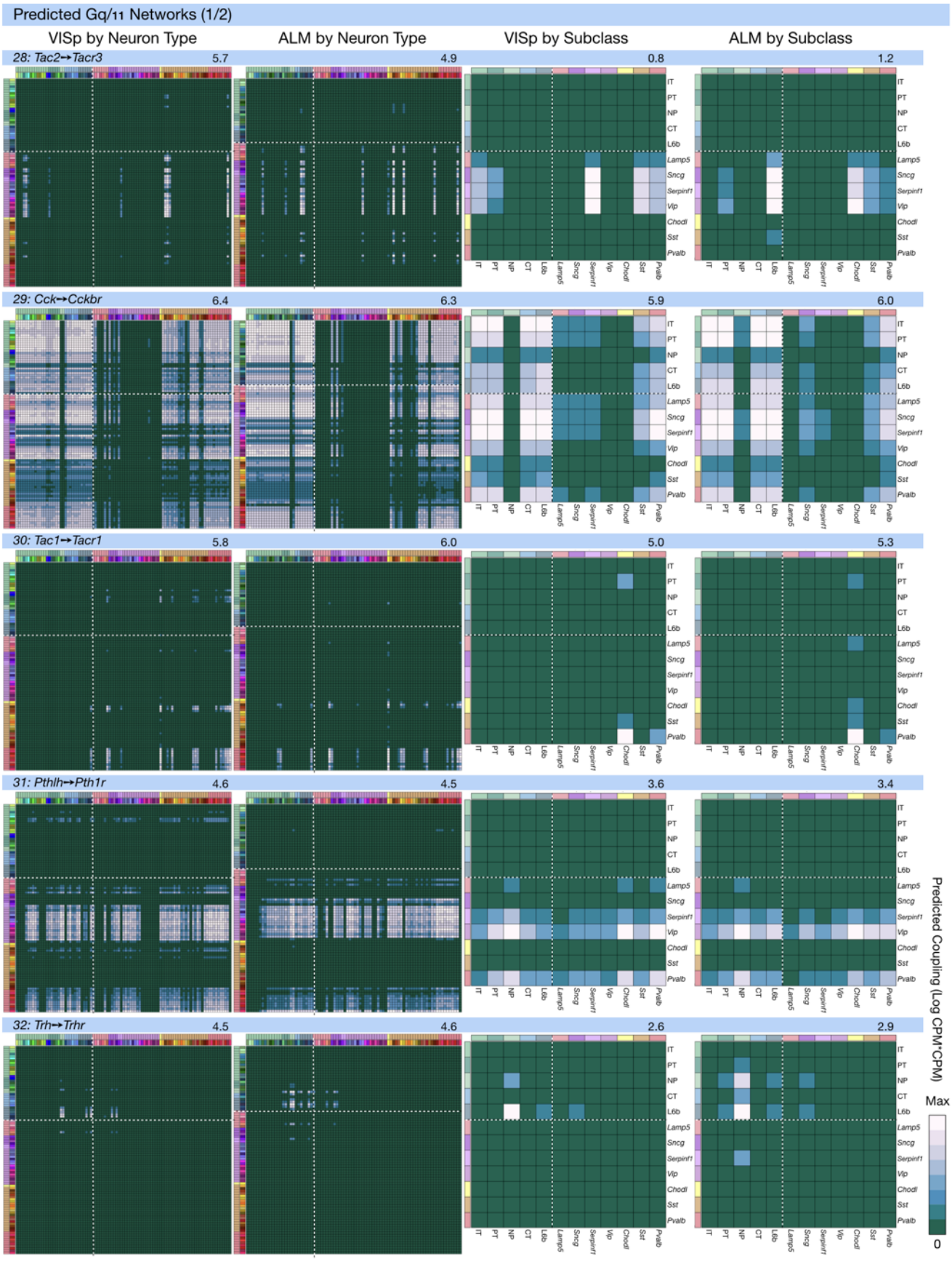


Figure 6 – figure supplement 1H (see legend, figure supplement 1A)

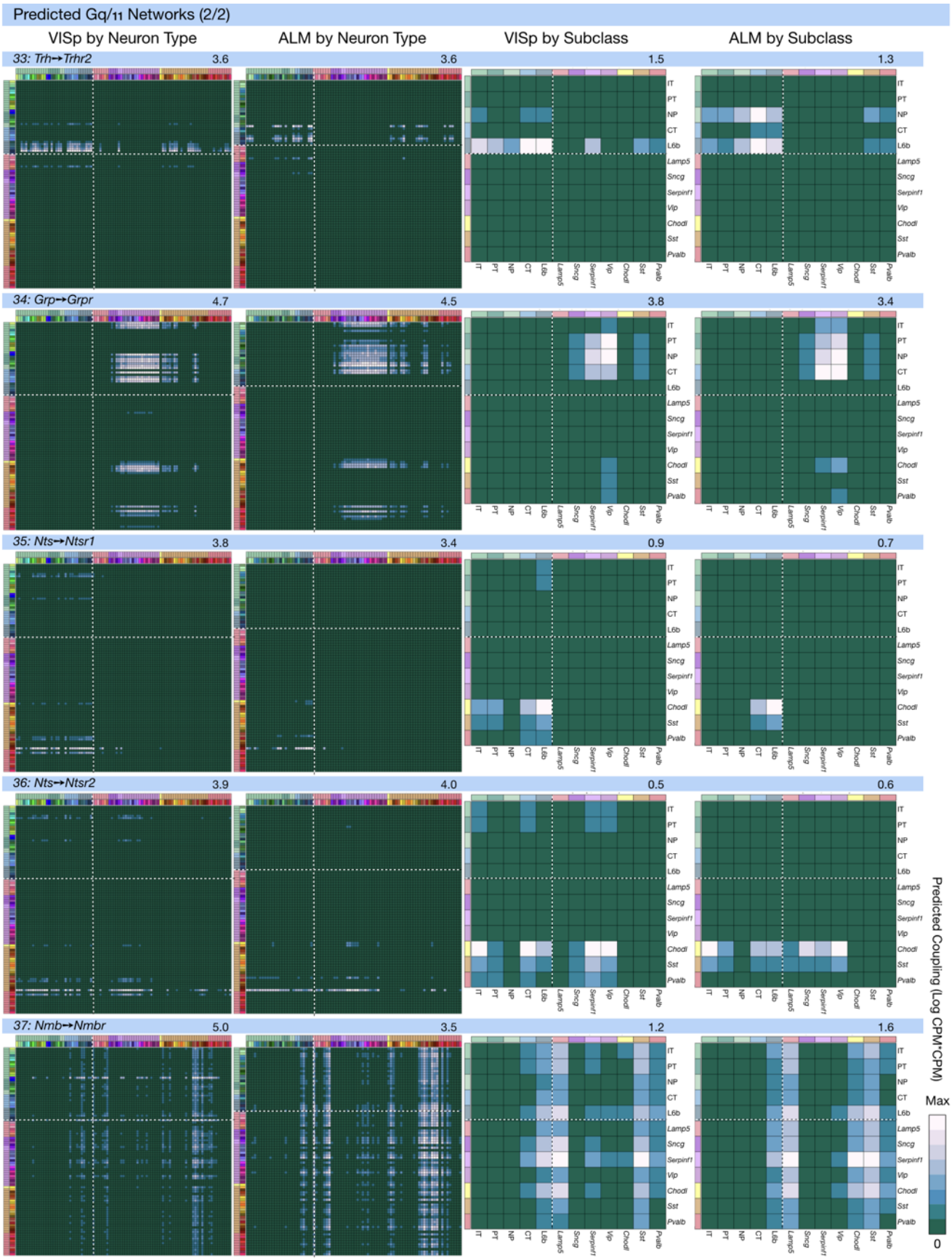
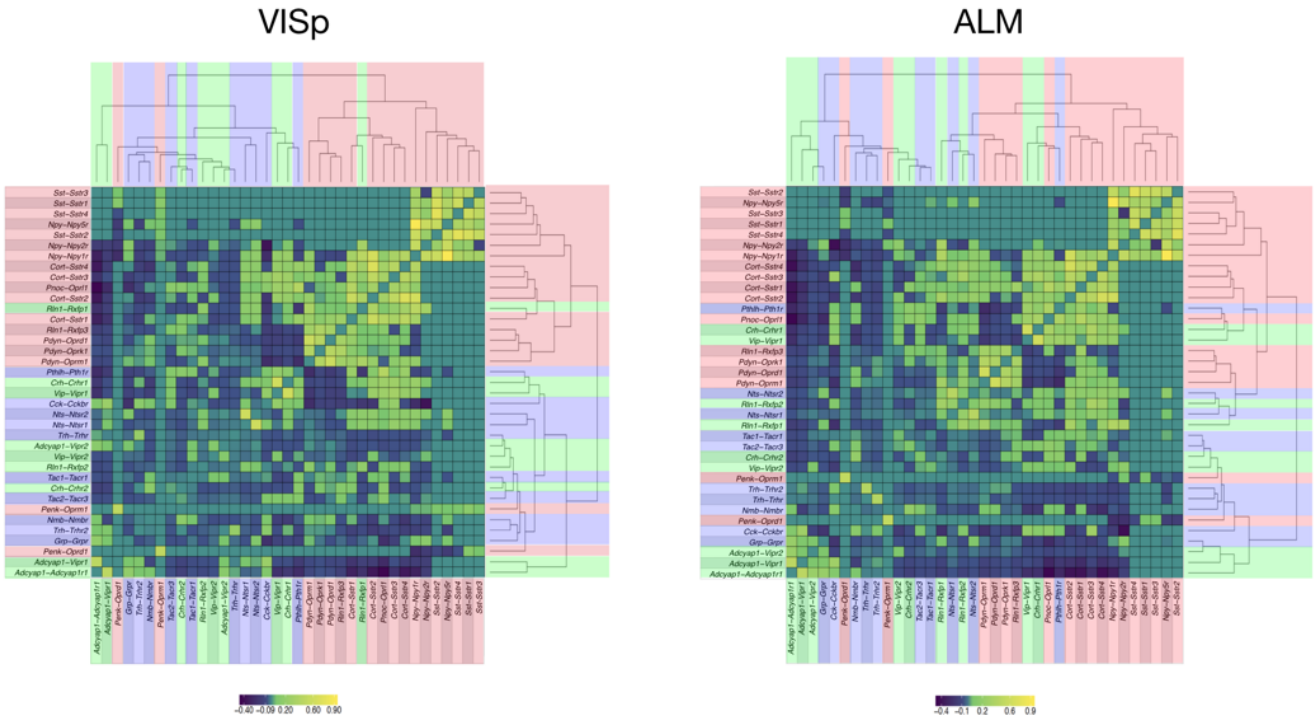
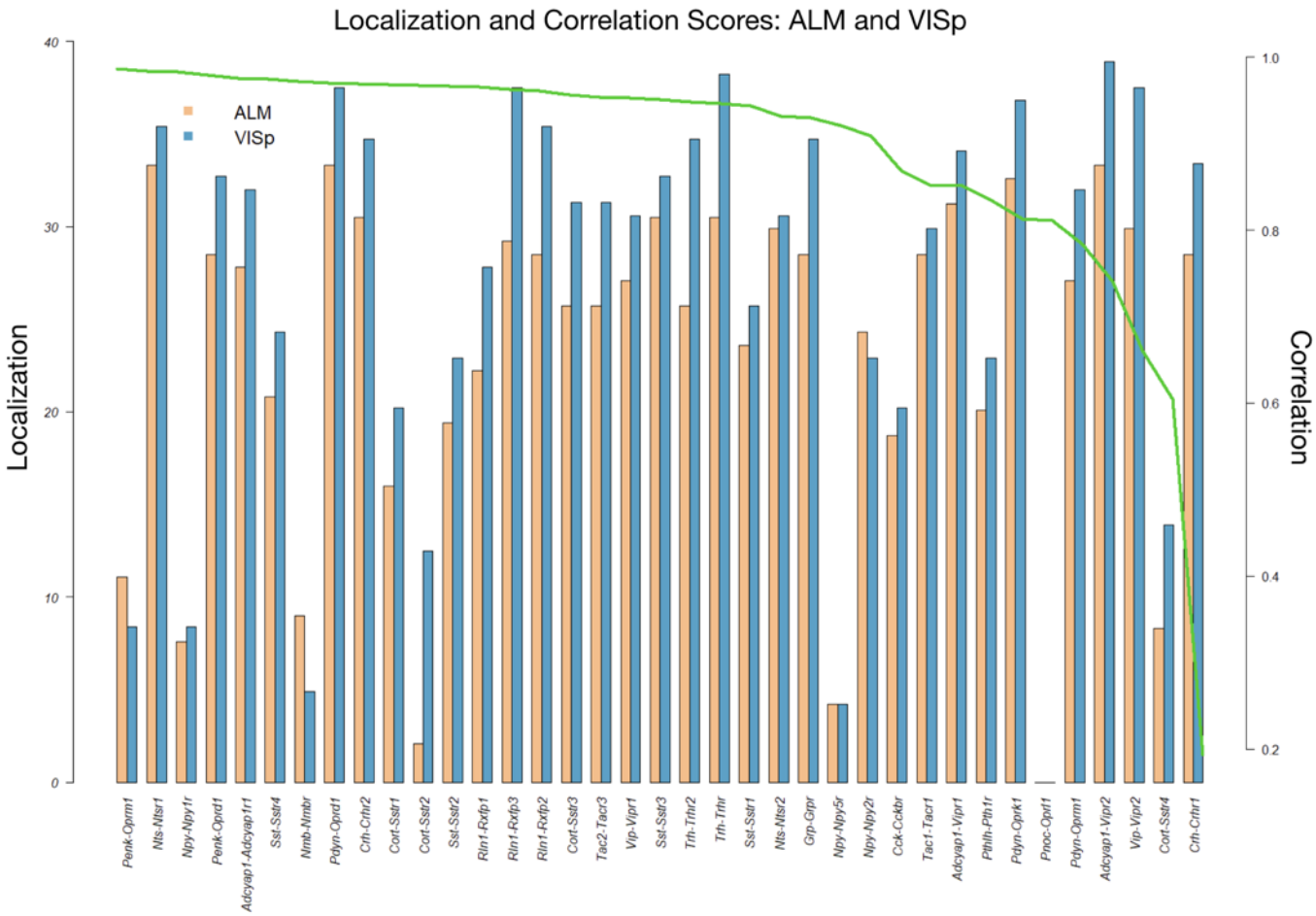


Figure 6 figure supplement 2



Hierarchical relationships of coupling matrices. For area VISp (left) and ALM (right), the NPP, NP-GPCR coupling matrices C are linearized to vectors and then clustered using hierarchical clustering (average linkage) of the Pearson correlation between associated NPP, NP-GPCR vector pairs. The result of clustering the correlation matrix demonstrates consistent organization between ALM and VISp, with Sst-Sstr1, Sst-Sstr2, Sst-Sstr3, Sst-Sstr4,Npy-Npy5r showing greatest similarity of pattern. A singular value decomposition (SVD) of the associated ALM, VISp matrices shows 20 eigenvalues are required to explain 95% of variance for ALM, and 18 eigenvalues to explain 95% variance for VISp. The complexity of these patterns suggests a highly non-redundant combinatorial code. Text fill colors indicate primary Gα family.

Figure 6 - figure supplement 3



Coupling matrix localization and correlation metrics for VISp and ALM areas. To measure the extent to which each (NPP, NP-GPCR) pair p demonstrates localized signal in its coupling matrix C_R^p in a region $R \in \{ALM, VISp\}$, we extract the top 10% percentile of C_R^p, then reduce the remaining expression signal to the pooled 12x12 major types matrix P_R^p. Let M be the maximum number of non-zero cells of P_R^t over all pairs t. Define the localization score L_R^p of the pair p as L_R^p = {M - # non-zero cells of P_R^p}. Another informative metric is the Pearson correlation $\rho(P_{ALM}^p, P_{VISp}^p)$ between the 12x12 major types matrices P_{ALM}^p and P_{VISp}^p. This figure shows the localization scores for each pair (NPP, NP-GPCR) for both ALM and VISp ordered by decreasing correlation $\rho(P_{ALM}^p, P_{VISp}^p)$. The result demonstrates a large number of localized pairs and localization that is generally independent of the correlation of the signal between ALM and VISp, suggesting reproducible localized signal across cortical regions.

Supplementary Files

Supplementary file 1. Gene ontology results for 18 NPP genes.
<Supplementary file 1.csv>

Supplementary file 2. Gene ontology results for 29 NP-GPCR genes.
<Supplementary file 2.csv>

Supplementary file 3. pFPKM expression data for 94 putative mouse NPP genes.
<Supplementary file 3.csv>

Supplementary file 4. Criteria for selecting 18 genes from a set of 39 cortically expressed NPP genes.
<Supplementary file 4.csv>

Supplementary file 5. pFPKM expression data for 84 putative mouse NP-GPCR genes.
<Supplementary file 5.csv>