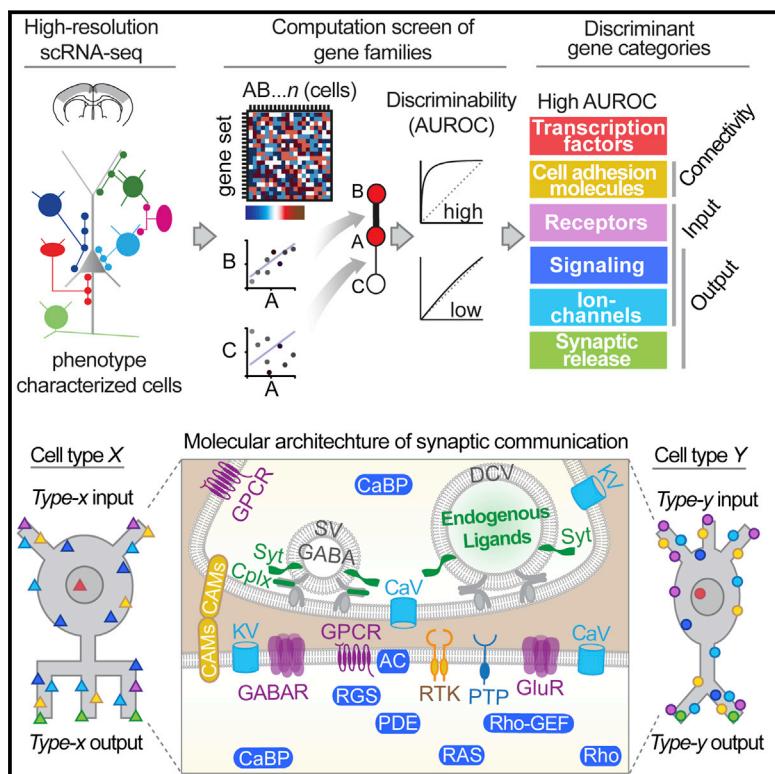


# Transcriptional Architecture of Synaptic Communication Delineates GABAergic Neuron Identity

## Graphical Abstract



## Authors

Anirban Paul, Megan Crow,  
Ricardo Raudales, Miao He, Jesse Gillis,  
Z. Josh Huang

## Correspondence

huangj@cshl.edu

## In Brief

GABAergic neuron types are distinguished by a transcriptional architecture that encodes their synaptic communication patterns.

## Highlights

- Single-cell transcriptome analysis of phenotype characterized GABAergic neurons
- Computation screen identifies gene families that distinguish GABA subpopulations
- 6 gene categories shape physiological input-output connectivity of GABA neurons
- Transcription profiles of synaptic communication encapsulate neuronal identity

# Transcriptional Architecture of Synaptic Communication Delineates GABAergic Neuron Identity

Anirban Paul,<sup>1</sup> Megan Crow,<sup>1</sup> Ricardo Raudales,<sup>1,2</sup> Miao He,<sup>1,3</sup> Jesse Gillis,<sup>1</sup> and Z. Josh Huang<sup>1,4,\*</sup>

<sup>1</sup>Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724, USA

<sup>2</sup>Program in Neuroscience, Stony Brook University, Stony Brook, NY 11790, USA

<sup>3</sup>Present address: Institutes of Brain Science, State Key Laboratory of Medical Neurobiology, Collaborative Innovation Center for Brain Science, Fudan University, Shanghai, China

<sup>4</sup>Lead Contact

\*Correspondence: huangj@cshl.edu

<http://dx.doi.org/10.1016/j.cell.2017.08.032>

## SUMMARY

Understanding the organizational logic of neural circuits requires deciphering the biological basis of neuronal diversity and identity, but there is no consensus on how neuron types should be defined. We analyzed single-cell transcriptomes of a set of anatomically and physiologically characterized cortical GABAergic neurons and conducted a computational genomic screen for transcriptional profiles that distinguish them from one another. We discovered that cardinal GABAergic neuron types are delineated by a transcriptional architecture that encodes their synaptic communication patterns. This architecture comprises 6 categories of ~40 gene families, including cell-adhesion molecules, transmitter-modulator receptors, ion channels, signaling proteins, neuropeptides and vesicular release components, and transcription factors. Combinatorial expression of select members across families shapes a multi-layered molecular scaffold along the cell membrane that may customize synaptic connectivity patterns and input-output signaling properties. This molecular genetic framework of neuronal identity integrates cell phenotypes along multiple axes and provides a foundation for discovering and classifying neuron types.

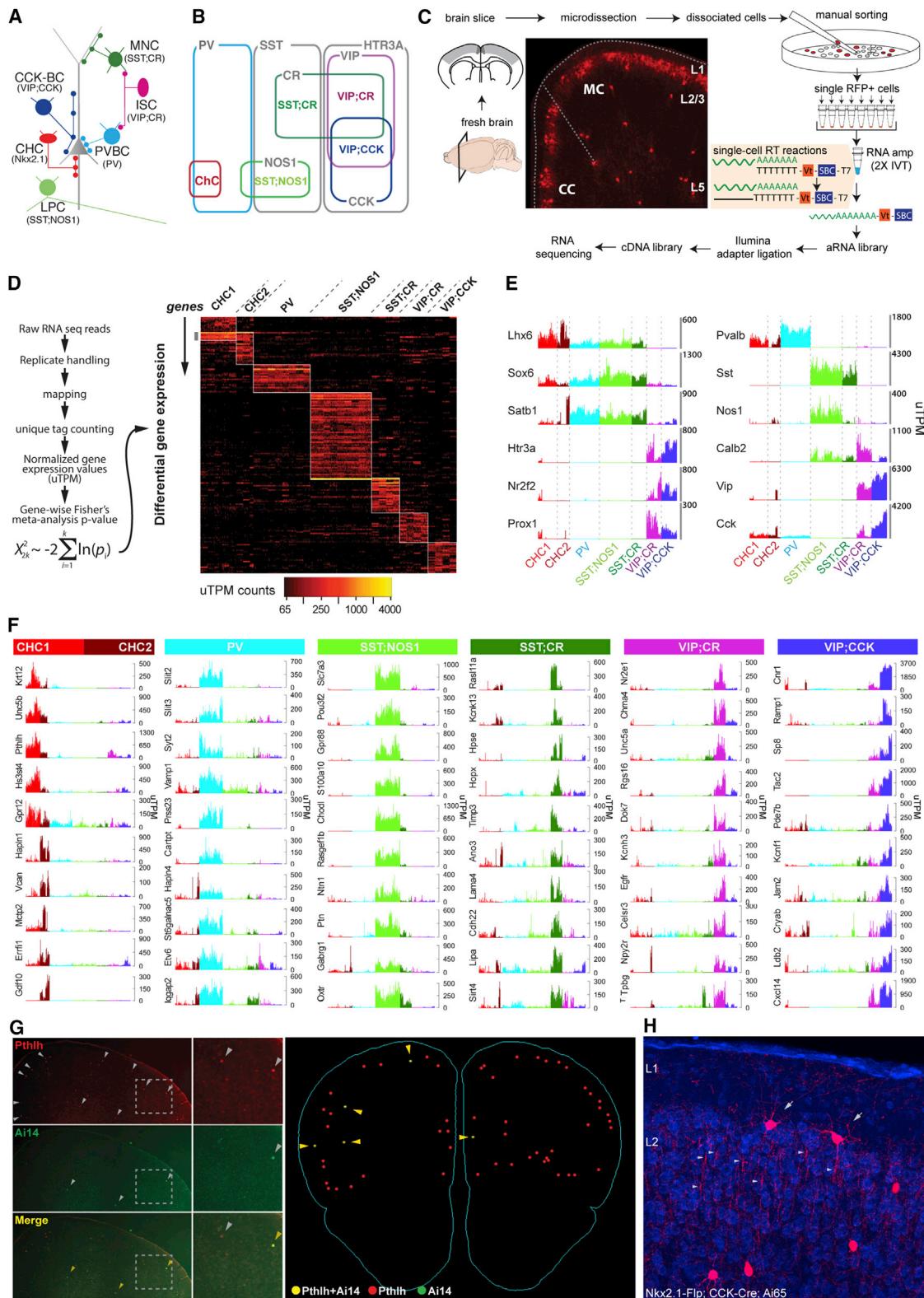
## INTRODUCTION

Since the discovery that individual neurons are the basic building blocks of the nervous system (Cajal, 1892), the diversity and heterogeneity of nerve cells have remained a challenge in deciphering the organization of neural circuits (Armañanzas and Ascoli, 2015). Recent technical advances in anatomical, physiological, and functional studies increasingly reveal multi-modal and multi-dimensional variations of neuronal phenotypes (Huang and Zeng, 2013). A fundamental question is whether these vari-

ations are largely subjective to measurements and only can be managed by operational grouping, or whether multiple distinct and congruent cell features can be integrated to define discrete “cell types” that reflect biological reality and mechanisms. The problem of neuronal diversity is unlikely to be solved without solving the equal, if not more fundamental, problem of neuronal identity that is the flip side of the cell-type coin (Seung and Sümbül, 2014). However, the biological basis of neuronal identity is poorly understood and cell definition and classification schemes in the brain remain contentious (DeFelipe et al., 2013).

As individual neurons constitute basic units of gene regulation, a major determinant of each neuron’s phenotype and function likely lies in its transcription programs. Recent studies classify neurons using high-throughput single-cell RNA sequencing (scRNA-seq) (Tasic et al., 2016; Zeisel et al., 2015). A major challenge is to map transcriptome-based statistical clusters, which are prone to technical noise and methodological bias, to the biological ground truth: their anatomical and physiological properties that constrain and contribute to cell function in neural circuits. In the retina, where cell types are among the best understood in the mammalian nervous system, scRNA-seq has identified distinct markers that correlate to known types and suggested novel types (Shekhar et al., 2016). In the cerebral cortex, where cell type definition is often ambiguous and controversial, scRNA-seq analyses have parsed multiple “transcriptional types” (Tasic et al., 2016; Zeisel et al., 2015), but their correlations to bona fide biological types jointly defined by anatomical and physiological features remain unclear. Thus although scRNA-seq allows comprehensive and quantitative measurements of gene expression, a fundamental unresolved issue is how transcription profiles might contribute to the molecular genetic basis of neuron types. Discovering the transcriptional basis of neuronal identity is prerequisite to deciphering neuronal diversity and enumerating cell census. Furthermore, high-resolution transcriptional portraits that mechanistically explain and predict the multi-faceted yet functionally congruent cell phenotypes have yet to be achieved.

Here, we have explored the transcriptional architecture underlying the core identity of GABAergic neurons in the cerebral cortex. Unlike recent studies that classify neurons using unsupervised statistical clustering of transcriptomes from unbiased

**Figure 1. Transcriptomic Analysis of GABAergic PCPs**

(A) Schematic of 6 PCPs with characteristic innervation patterns.

(B) Molecular markers parsing cortical GABA neurons into 3 non-overlapping populations and 6 PCPs.

(legend continued on next page)

(Zeisel et al., 2015) or relatively broad populations (Tasic et al., 2016), we analyzed single-cell transcriptomes of 6 genetically labeled and phenotypically well-characterized GABAergic types or subpopulations to discover their core molecular features. Using these anatomically and physiologically defined subpopulations as an assay, we designed a computational genomics strategy to screen through the ~620 HGNC (Human Genome Nomenclature Committee) gene families for those whose differential expression reliably distinguish these subpopulations. Remarkably, approximately 40 gene families implicated in regulating synaptic communication best distinguish these subpopulations. These gene families constitute 6 functional categories that include cell-adhesion molecules, neurotransmitter and modulator receptors, ion channels, regulatory components of membrane-proximal signaling pathways, neuropeptides and vesicular release components, and transcription factors. Combinatorial and coordinated expression of select family members across these categories shapes a molecular scaffold along the cell membrane that appears to customize the pattern and property of synaptic communication for each cell population. We further provide evidence that transcription factor profiles register the developmental history of GABAergic neurons and contribute to concerted gene expression patterns that shape cell phenotypes. These findings suggest that neuronal identity is encoded in a transcriptional architecture that orchestrates functionally congruent expression across multiple gene families to customize the patterns and properties of their input-output communication.

## RESULTS

### Single-Cell Transcriptomes of Phenotype-Characterized GABAergic Cells

Our overall strategy is to compare high-resolution transcription profiles of a set of well-characterized GABAergic neurons defined by their anatomical, physiological, and developmental attributes. We developed combinatorial recombinase driver lines to capture 6 GABAergic subpopulations (He et al., 2016) (Figures 1A and 1B; see the STAR Methods): (1) the *Nkx2.1-CreER* driver allows lineage and birth timing based targeting of chandelier cells (CHCs) that target the axon initial segment (AIS) of pyramidal neurons (Taniguchi et al., 2013); (2) the *PV-Cre* driver labels fast-spiking basket cells (PVBCs) that innervate the perisomatic region (Hu et al., 2014); (3) the *SST-Flp;NOS1-CreER* drivers target a unique type of long-projecting GABAergic cells (LPCs) (Kilduff et al., 2011); (4) the *SST-Flp;CR-Cre* drivers include Martinotti cells (MNCs) that target distal dendrites (Silberberg and Markram, 2007) and likely another cell type; (5) the *VIP-Flp;CR-Cre* drivers include interneuron-selective cells (ISC) (Staiger

et al., 2004) and likely other types; and (6) the *VIP-Flp;CCK-Cre* drivers include CCK basket cells (CCKCs) (Armstrong and Soltesz, 2012) and likely other types. Accumulated anatomical, physiological, and molecular evidence indicate that these are non-overlapping subpopulations, and CHCs, LPCs, and PVBCs are considered cardinal types (He et al., 2016). We define these 6 populations as phenotype characterized populations, or PCPs.

Using manual sorting of single red fluorescent protein (RFP)-labeled cells from microdissected motor and somatosensory cortex of 6-week-old mice (Figure 1C), we obtained a high-resolution transcriptome dataset of ~584 cells (CHC1 = 80, CHC2 = 52, PVBCs = 127, LPCs = 136, MNCs = 62, ISC = 63, and CCKCs = 64 cells) from the 6 PCPs (Figure S1D; STAR Methods). Compared with previous methods (Shekhar et al., 2016; Tasic et al., 2016; Zeisel et al., 2015), our linear amplification with 10-bp unique molecular identifiers (UMIs) achieved more comprehensive and quantitative transcriptome measurements (detecting ~9,000 genes/cell; Figure S1H; STAR Methods); this facilitated more in-depth analysis of molecular profiles that contribute to the phenotypes and identity of PCPs.

Our analysis revealed 190 genes differentially expressed (DE) among PCPs with each gene expressed at >50 uTPM (unique transcript per million), >4-fold enrichment, and with a p value <  $5 \times 10^{-4}$  (Figure 1D; Table S1). We confirmed known markers for medial ganglionic eminence (MGE)- and caudal ganglionic eminence (CGE)-derived interneurons and the 6 PCPs (Figure 1E). We further profiled CHCs from upper (L2/3, CHC1) and deeper (L5/6, CHC2) layers and identified ~11 genes enriched in CHC2 over CHC1 (Figure 1F; Table S1). We validated PCP-specific expression of ~10 transcripts by mRNA in situ hybridization (Figures 1G and S2A). We discovered a putative pan-CHC transcript Pthlh: ~95% of RFP-labeled CHCs were positive for Pthlh (136/143 cells), and their laminar distribution recapitulates CHC pattern in frontal cortex (Figures 1G and S2A). To validate CCK expression in CHCs (Figure 1E), we demonstrated that an intersection of *Nkx2.1-Flp* and CCK-Cre labeled CHCs at the L1, L2 border (Figure 1H).

### A Computation Genomics Screen Identifies Gene Families and Categories that Distinguish PCPs

Cellular properties emerge from operations of macromolecular machineries comprising interacting components, each often implemented as one of multiple variants encoded by a gene family. Thus variations of cell properties often result from differential expression of select gene family members with characteristic biochemical and biophysical properties that confer customized features to cellular machines (Hartwell et al., 1999). We hypothesized that phenotypic differences among PCPs

(C) Experimental workflow.

(D) Bioinformatics pipeline (left) and differential expression genes across PCPs (right).

(E) Validation of known PCP markers; uTPM, unique transcripts per million.

(F) Novel PCP markers.

(G) Pthlh mRNA (gray arrowheads) co-localizes with RFP-labeled CHCs (yellow arrowheads) shown by fluorescence in situ hybridization (FISH) (left). Serial 3D reconstruction shows >95% RFP cells are Pthlh<sup>+</sup> (right).

(H) Cck<sup>+</sup> CHCs are labeled in *Nkx2.1-Flp;CCK-Cre;Ai65* cortex.

Red, RFP; blue, DAPI; arrow, CHC; arrowhead, CHC axon boutons.

See also Figures S1 and S2, Table S1, and the STAR Methods.

result from systematic differential transcription across multiple gene families. Our experimental design, whereby single-cell transcriptomes derive from 6 PCPs, provided an effective assay to systematically screen for such gene ensembles that distinguish PCPs.

The essence of our computational screen, driven by a supervised, machine-learning-based algorithm, MetaNeighbor (Crow et al., 2017), is to detect whether a given set of genes shows correlated expression among cells of the same identity (Figure 2A). As our single-cell transcriptomes derive from 6 PCPs, this data structure allowed characterizing the similarity between all cell pairs using co-variation of expression level in many known gene sets and measure whether a given set correctly links cells of known identity. In network formalism, cells are linked as probabilistically related based on the similarity of their transcriptional profiles across a given set of genes (Figure 2A). This network classifies cells based on their proximity: closely linked neighbors are predicted to share an identity (see the STAR Methods). A subset of PCP labels is first applied to cells, giving a subnetwork with known identities that can classify unlabeled cells. We then hold back the PCP identity of some cells and attempt to predict their identities using this subnetwork of known identities. The efficacy of this test, reported as mean area under the receiver operator characteristic curve (AUROC), maps to the probability of a correct assignment when making a single binary choice (Figure 2A). Having constructed a computational assay for cell identity, we vary the transcriptomic features (e.g., gene families) used to characterize cells as neighbors. This computation screen thus selects gene ensemble features which jointly distinguish cell identities.

We first screened for gene ensembles according to gene ontology (GO) terms, using randomized labels (AUROC~0.5) as controls. GO terms containing the keyword “synaptic” gave the highest AUROC score (0.91–0.98), suggesting that genes implicated in synaptic connectivity and function are most discriminating for PCPs (Figure 2B; Table S2). To identify more specific gene categories, we screened through all ~620 gene families annotated in the HGNC database (see the STAR Methods). We identified ~40 gene families (i.e., 7% of all gene families) with AUROC scores >0.75, generally regarded as a stringent threshold (Figure 2C; Tables S3 and S4). Strikingly, these gene families all fell into only 6 functional categories (Figures 2C and 2D): (1) cell-adhesion molecules, (2) receptors for neurotransmitters and modulators, (3) voltage-gated ion channels, (4) regulatory signalling proteins, (5) neuropeptides and vesicle release machinery, and (6) transcription factors. It is immediately evident from this list that, except for transcription factors (TFs), all gene categories encode proteins that localize along or close to cell and synaptic membrane (Figure 2D) and contribute to a singular aspect of neuronal biology—synaptic communication—which is implemented through synaptic connectivity and input-output signaling properties (Figure 2E).

To validate this discovery, we applied MetaNeighbor to two independent scRNA-seq datasets (Tasic et al., 2016; Zeisel et al., 2015). Despite differences in experimental design and methods our meta-analysis yielded high replicability on PCP-distinguishing gene families (Figure 2F). Nearly identical gene families also discriminated the 6 PCP-equivalent cell populations in all three

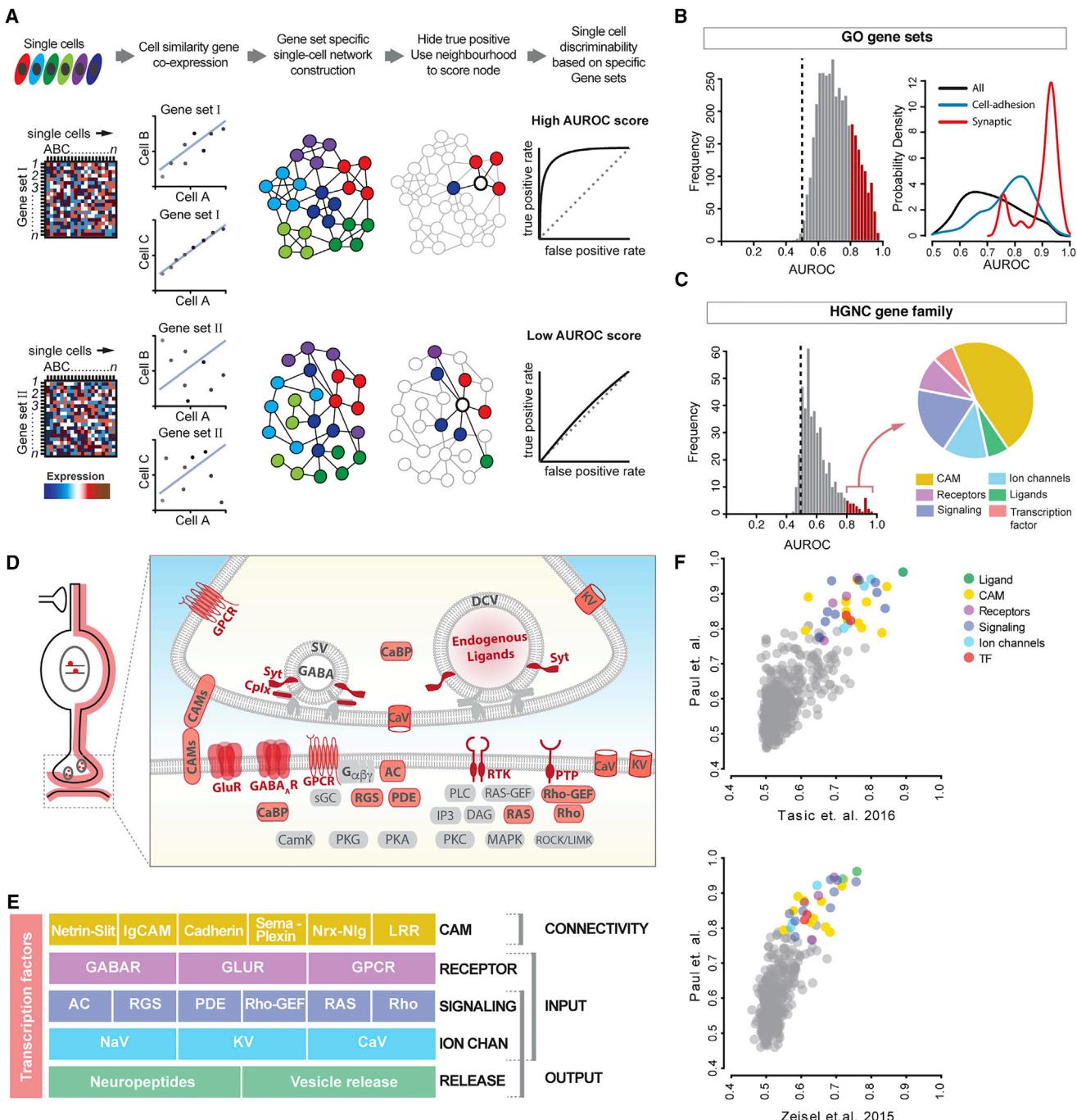
datasets, and the AUROC values of these families (Figure 2C) were well correlated in pair wise comparisons even though the scores from the other two datasets were modestly lower (Figure 2F).

We then examined gene families that were not predictive of PCP identity. ~243 families with AUROC <0.6 (Table S5) were mostly implicated in generic cell functions ranging from DNA replication to energy conversion, etc. This analysis also revealed gene families that a priori might be assumed to predict PCP identity but did not. These include (1) a substantial set of cytoplasmic signaling kinases (e.g., the entire mitogen-activated kinase cascade), phospholipases (30 members); (2) most cytoskeleton components (actins, myosin, dyneins, kinesins); and (3) core components of vesicle release machinery (Table S5). These non-predictive gene families contrasted with the top 40 predictive families in two major properties: they are all cytosolic instead of membrane or membrane proximal, and they often make up the generic scaffold of macromolecular complexes or pathways instead of the key regulatory components.

### Differential Expression of Cell-Adhesion Molecules and Carbohydrate-Modifying Enzymes Suggests a Large Capacity for Cell-Surface and Extracellular Matrix Labels

Each GABAergic neuron receives inputs from and extends output to diverse pre- and post-synaptic neurons, respectively (Figure 3A). The cell-adhesion molecules (CAMs) that regulate their morphology, connectivity, synaptic transmission and plasticity are largely unknown. Our computational screen identified multiple CAM families that effectively discriminate PCPs (Figures 3B–3E). Based on broadly annotated HGNC families and the literature (de Wit and Ghosh, 2016; Kolodkin and Tessier-Lavigne, 2011; Takahashi and Craig, 2013), we selected ~275 genes encoding all major neuronal CAMs and organized them into 12 adhesion groups according to sequence homology and receptor-ligand relationships (Figure 3B; Table S4). Nearly all major groups of neuronal CAMs implicated in different aspects of neural development are expressed in PCPs, and each PCP expresses ~200 CAM genes (Figure 3C). This was an underestimate of CAM diversity as our RNA-seq method does not detect splicing variants. Among the ~275 CAM genes, 130 show highly distinct subpopulation profiles (Figure 3E; Table S6). Strikingly, multiple CAM families each manifest differential expression among PCPs (Figures 3F and S3A–S3D). For example, UNC5 members and their ligand netrin1 are differentially expressed; UNC5b is highly specific to CHCs (Figures S2B and S3C); these receptor-ligand pairs might mediate cell-cell recognition (Figure S3A).

Each of the major synaptic adhesion families (e.g., neuroligin, neuroligin, protein tyrosine phosphatases, leucine-rich repeat proteins, and Slitrks) is differentially expressed among PCPs (Figures 3D–3F and S3E). Cell-specific expression of LRRs might contribute to post- and trans-synaptic specializations that customize the property of synapse types defined by pre- and post-synaptic neuron identities. Together, cell- and synaptic adhesion families likely constitute a comprehensive mosaic of multi-faceted cell-surface code throughout the neuronal membrane.

**Figure 2. Gene Families and Categories that Distinguish PCPs**

(A) MetaNeighbor schematic. scRNA-seq values for gene sets are used to construct cell networks such that cells similar in gene expression space are close neighbors (connected by lines). PCP identity labels (colors) are then withheld, and their identities are inferred based on connectivity to immediate neighbors. The probability of being identified as the correct PCP is reported as an AUROC score (0.5 is at chance).

(B) Left: AUROC value distribution of ~3,800 GO terms. Red, AUROC >0.8. Right: GO term probability density by keyword; “synaptic” and “cell-adhesion” are skewed with AUROC >0.8.

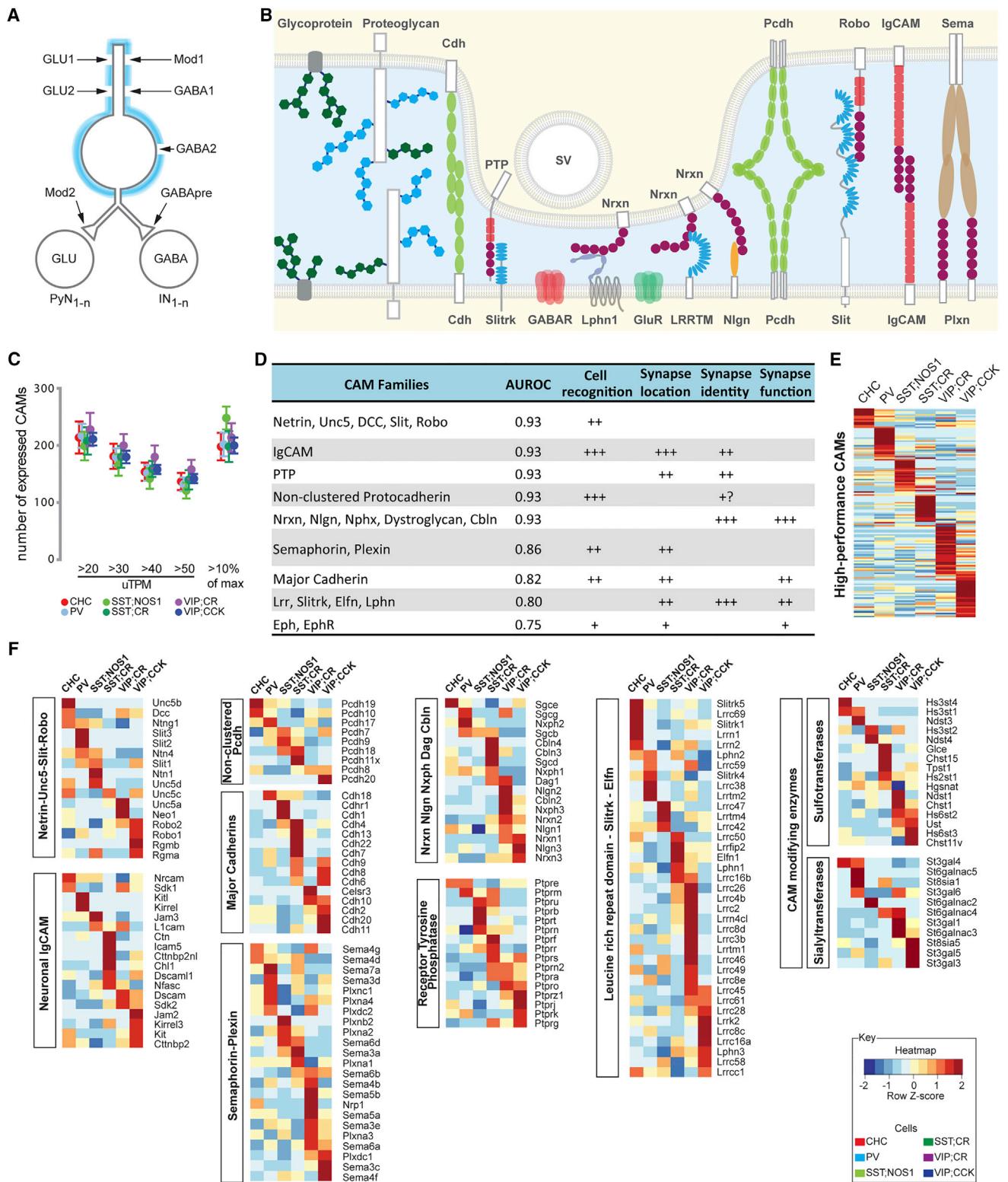
(C) AUROC distribution of 442 HGNC gene families. ~40 families (red bars) in 6 categories (pie chart) are highly predictive of PCP identities (AUROC  $\geq 0.8$ ).

(D) Schematic showing that high-performance gene families (except TFs) encode proteins that primarily localize along cell and synaptic membrane.

(E) High-performance gene families constitute 5 layers of functional categories that organize synaptic connectivity and I/O signaling.

(F) MetaNeighbor analysis of two independent scRNA-seq datasets yields similar rank order of gene families.

See also Figure S2, Tables S2, S3, S4, and S7, and Document S1.



**Figure 3. Differential Expression of Cell-Adhesion Molecules and Carbohydrate-Modifying Enzymes among PCPs**

(A) A single GABAergic neuron receives multiple sources of glutamatergic, GABAergic, and modulatory (Mod) inputs and innervates large sets of pyramidal neurons (PyN) and interneurons (IN).

(B) Multiple families of CAMs and glycoproteins provide extracellular coating, cell-surface, and synaptic labels.

(legend continued on next page)

We further discovered prominent differential expression in two families of carbohydrate modifying enzymes, sulfotransferases (AUROC = 0.88) and sialyltransferases (AUROC = 0.85), that may increase the molecular diversity of glycosylated CAMs and proteoglycans on the cell membrane and in extracellular matrix (Figures 3F and S3D). This suggests that each PCP might produce a characteristic cell coat through distinct carbohydrate modification patterns to diversify proteoglycans that facilitates or prevents cell interaction at a distance.

### Differential Expression of Transmitter and Modulator Receptors Shapes the Input Properties of PCPs

#### *Ionotropic Glutamate Receptors*

Ionotropic glutamate receptors (iGluRs) play key roles in excitatory synaptic signaling and plasticity. Glutamatergic synapses in GABAergic interneurons often contain higher proportions of CP-AMPARs ( $\alpha$ -amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid receptor) and GluN2B-NMDARs, although the ratio between the two types of AMPARs and NMDARs vary significantly among different cell populations (Akgül and McBain, 2016). Substantiating previous results from hippocampal interneurons (Akgül and McBain, 2016), we found that the mRNA levels and relative ratio of CP- versus Cl-AMPAR subunits in PCPs vary in a highly cell-type-dependent pattern (Figures 4A–4D and S4A). These results suggest cell-type-dependent composition and correlation of AMPA and NMDA receptor pore-subunits, especially with regard to the relative abundance and ratio of CP- versus Cl-AMPARs and 2B- versus 2A-NMDARs.

In addition to pore-forming subunits, native AMPARs incorporate multiple auxiliary subunits that regulate AMPAR membrane trafficking, synaptic targeting, gating, and signaling (Haering et al., 2014). TARP, SHISA, and CNIH family auxiliary subunits show striking cell-specific expression patterns (Figures 4C–4E). TARP $\gamma$ 2 is enriched in PV cells, TARP $\gamma$ 3,  $\gamma$ 8, and SHISA6 are enriched in SST/CR cells, TARP $\gamma$ 3 and SHISA9 are enriched in VIP/CCKCs. While PV cells predominantly express one auxiliary subunit (TARP $\gamma$ 2), SST/CR cells express at least 6 types (TARP $\gamma$ 2,  $\gamma$ 3,  $\gamma$ 8,  $\gamma$ 5,  $\gamma$ 7, and SHISA6). Whereas pore-subunits differ in expression levels, auxiliary subunits often show on/off expression among PCPs (Figure 4E). Thus different GABAergic neurons may assemble a specific set of native AMPARs with distinct pore and auxiliary subunit compositions, post-synaptic distribution patterns and biophysical properties. This large repertoire of native AMPARs may achieve cell-type- and synapse-specific transmission and plasticity of glutamatergic inputs according to different pre-synaptic sources.

#### *Ionotropic GABA Receptors*

GABA receptors (GABA<sub>A</sub>Rs) mediate fast inhibitory neurotransmission and are assembled as heteropentameric chloride chan-

nels from 19 subunits, typically consisting of 2 $\alpha$ , 2 $\beta$ , and 1 $\gamma$  subunits (Olsen and Sieghart, 2008) (Figures 4G–4H). The vast majority of possible subunit combinations remain tentative as previous studies do not achieve cellular resolution of subunit co-expression. Whereas  $\gamma$ 2 is regarded as the ubiquitous obligatory subunit of most if not all synaptic GABA<sub>A</sub>Rs that mediate phasic inhibition,  $\gamma$ 3 is sparsely expressed in cortical neurons of unknown identity and can assemble with  $\alpha$  and  $\beta$  to form synaptic receptors with slowly decaying inhibitory post synaptic currents (IPSCs) (Kerti-Szigeti et al., 2014). We found that, surprisingly,  $\gamma$ 3 is not only prevalent but also transcribed at much higher levels than  $\gamma$ 2 subunits in all 6 PCPs (Figure 4H). This suggests that  $\gamma$ 3 might contribute to the assembly of a class of slower decaying, longer duration synaptic GABA<sub>A</sub>Rs in GABAergic neurons. Different PCPs show specific subunit profiles and levels (Figures 4G and 4H). PVBCs express the largest variety (all except  $\alpha$ 2,  $\alpha$ 6, and  $\gamma$ 1) and overall highest levels of subunits, and uniquely high level of the GABA<sub>A</sub>R clustering/scaffolding protein gephyrin. In contrast, SST/CR cells express the least variety (mainly  $\alpha$ 3,  $\beta$ 1/3, and  $\gamma$ 2/3) and lowest overall levels. SST/NOS1 cells are distinguished by predominant expression of slow kinetics  $\alpha$ 2-containing GABA<sub>A</sub>Rs and, surprisingly, the exceedingly rare  $\gamma$ 1 subunit which is thought to assemble extra- or non-synaptic GABA<sub>A</sub>Rs (Dixon et al., 2014). PVBCs and CHCs express the  $\delta$  subunit, known to assemble extra-synaptic GABA<sub>A</sub>Rs that possibly localize to pre-synaptic terminals.

These cell resolution profiles, when considered with the well-characterized connectivity patterns among PCPs, suggest that distinct GABA<sub>A</sub>R subtypes with specific subunit combinations are likely targeted to specific connections that match the pre-synaptic terminals to optimize inhibitory transmission properties (Figure S4B). Cell-type-specific subunit expression may assemble GABA<sub>A</sub>R subtypes with distinct biophysical and pharmacological properties and subcellular localization and thus might customize inhibitory transmission between cell types.

#### *Neuromodulatory and G-Protein-Coupled Receptors*

Cortical GABAergic neurons received a range of modulatory inputs that convey diverse signals of brain states through a large family of G-protein-coupled receptors (GPCRs). We found that whereas MGE-derived PCPs are characterized by higher levels and larger variety of iGluRs and GABA<sub>A</sub>Rs, CGE-derived PCPs express much larger variety of neuromodulatory receptors (Figures 4I–4L). For example, although PV and VIP/CCKCs both innervate the perisomatic regions of pyramidal neurons, PVBCs enrich for only a few modulatory receptors (e.g., CCK2R, Oprd1), whereas VIP/CCKCs express multiple GPCRs for serotonin, acetylcholine, norepinephrine, endocannabinoid, adrenaline, NPY and VIP (Figure 4J–4L). Considered with their iGluR and GABA<sub>A</sub>R profiles, the results suggest that, similar to their hippocampal homologs (Armstrong and Soltesz, 2012), cortical

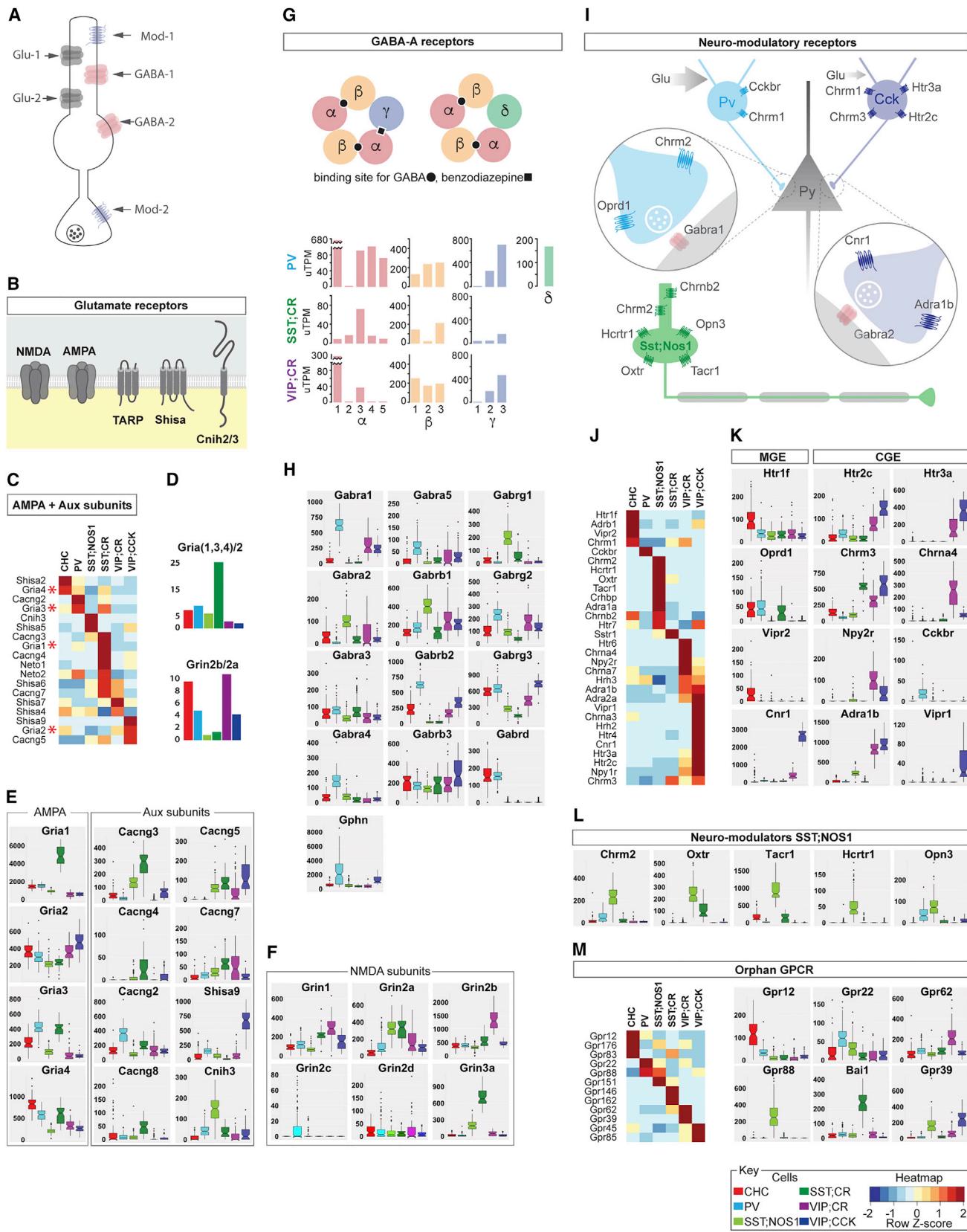
(C) ~200 different CAM genes are expressed in each PCP estimated using sliding expression values or 10% of maximum expression value as thresholds. Error bars show SD.

(D) Major ligand-receptor cell-adhesion systems and their roles in synaptic connectivity; all are highly discriminative of PCPs. "+" denotes the degree of involvement in the listed function.

(E) Differential expression (DE) of 136 CAM genes across 6 PCPs.

(F) Differential expression of 8 cell-adhesion systems and 2 carbohydrate-modifying enzymes families among PCPs.

See also Figure S3, Tables S5 and S6, Document S1, and STAR Methods, Notched boxplots.



(legend on next page)

PVBCs are recruited by fast and precise excitatory and inhibitory inputs from local cortical sources, whereas CCKBCs are profoundly modulated by subcortical inputs that represent internal drive and behavioral states (Table S7).

As an exception among MGE-derived neurons, the long projection SST/NOS1 cells show lower levels of iGluRs and extra- or non-synaptic  $\gamma 1$ -containing GABA<sub>A</sub>Rs but express a large and unusual set of modulatory receptors including hypocretin, oxytocin, neurokinin, Tacr1 (Figure 4L), which are released from hypothalamic centers that regulate global brain states (Kilduff et al., 2011). These results depict a cell type with weak phasic excitatory and inhibitory inputs but a wide range of tonic subcortical modulatory inputs, consistent with its activation by homeostatic sleep drive and speculated role in regulating global cortical networks (Kilduff et al., 2011) (Table S7).

PCPs are further characterized by their expression of orphan GPCRs for unknown or unproven ligands. Each PCP can be distinguished from all other by specific expression of at least 2 orphan GPCRs (Figure 4M). In particular, the metabotropic zinc ( $Zn^{2+}$ ) sensor GPR39/mZnR is specifically expressed in VIP/CCK and to a less extent SST/NOS1 cells. Upon zinc binding, GPR39 promotes KCC2 membrane trafficking, thereby enhancing GABA<sub>A</sub>R mediated hyperpolarization (Chorin et al., 2011). Thus GPR39 in VIP/CCKCs might mediate activity-dependent modulation of their excitability.

#### Differential Expression of Voltage-Gated Ion Channels and Electrophysiological Properties of PCPs

The ion homeostasis and sophisticated intrinsic and synaptic physiology properties are shaped by several families of voltage-gated ion channels (VGICs), each contains diverse family members with characteristic biophysical properties (Yu and Catterall, 2004). We demonstrate extensive differential transcription profiles within and across multiple VGICs families among PCPs (Figure S4). Within the Nav and Cav family, major pore-forming subunits are broadly expressed among PCPs, often with different expression levels (Figures S4C–S4E). Cav auxiliary subunits ( $\beta 1$ -2,  $\alpha 2$ , and  $\delta 1$ -4) show more distinct, often binary pattern (Figure S4D), suggesting cell-specific regulation of the trafficking, gating, and kinetics of pore forming subunits. Within

the Kv family, different subsets are differentially enriched among PCPs (Figure S4C); there is often a tight correlation between expression of principle subunits (e.g., KcnA1/Kv1.1 and KcnA2/Kv1.2) and their matching auxiliary subunits (Kv $\beta$ 1-3, Kcnab1, b2, and b3) in specific PCPs (e.g., PVBCs) (Figure S4), implying cell-specific assembly of functional channel complex. These results suggest that differential and correlated expression across multiple families of VGIC subunits may customize electrical signaling among PCPs (Figures S4C–S4G).

#### Differential Expression of Signaling Proteins in Calcium, Cyclic Nucleotide, and Small Guanosine Triphosphate Second Messenger Pathways in PCPs

As a conserved cell signaling scheme, a large repertoire of surface receptors transduce diverse extracellular signals into a small set of intracellular second messengers, which trigger enzyme cascades that regulate excitability, transmitter release, metabolism, neurite motility and gene expression (Alberts et al., 2014) (Figure 5A). Superimposed upon these conserved second messenger cascades, different cell types deploy a large set of regulatory signaling proteins to control the spatiotemporal dynamics of each signal transduction pathway to achieve appropriate responses. We discovered that whereas most kinase cascades and signal proteins are broadly expressed, a small set of regulatory protein families in the calcium, cyclic nucleotide, and small guanosine triphosphate (GTPase) pathways are highly differential among PCPs and may tailor specific signal transduction properties (Figure 5A).

#### $Ca^{2+}$ Binding Proteins

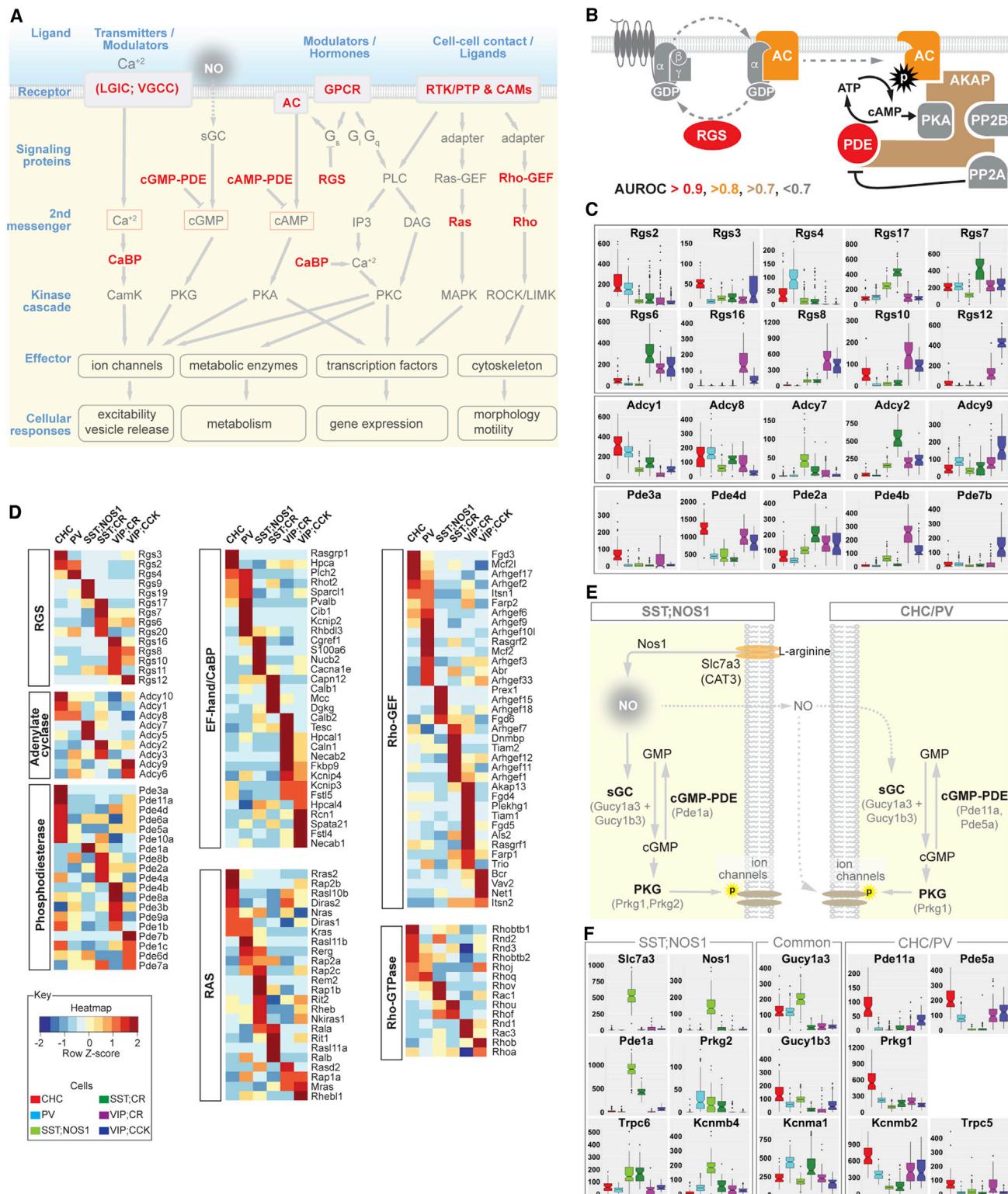
Each PCP expresses a set of ~5–8 different  $Ca^{2+}$ -binding proteins (CaBPs) (Figures 5D and S5E), many of which are in fact signaling proteins (e.g., Rasgrp1 in CHCs). This result suggests that differential expression of multiple  $Ca^{2+}$  binding and signaling proteins might shape the spatiotemporal dynamics and specificity of  $Ca^{2+}$ signaling among PCPs.

#### Adenylyl Cyclase, Phosphodiesterase, and Cyclic AMP Signaling

In mediating GPCR signaling, the synthesis, degradation, and spatiotemporal dynamics of cyclic AMP (cAMP) are stringently regulated at each step (Halls and Cooper, 2011). We found

**Figure 4. Differential Expression of Transmitter and Modulatory Receptors among PCPs**

- (A) Schematic of transmitter and modulatory receptors on a generic GABAergic neuron.
  - (B) Schematic of glutamate receptor core subunits and auxiliary proteins that form native receptors.
  - (C) Differential expression of AMPAR core subunits and auxiliary proteins across PCPs; SST;CR cells express the greatest diversity of AMPARs.
  - (D) Top: SST;CR cells show highest Grin1,3,4 (GluN1,A3A4)/Grin2 (GluN2) ratio among PCPs. Bottom: Most GABAergic neurons have more Grin2b (GluN2B) than Grin2a (GluN2A) receptors; the reverse is true in SST neurons.
  - (E) AMPAR core and auxiliary subunits shows striking differences among PCPs.
  - (F) Differential expression of NMDAR subunits; glycine-activated Grin3a (GluN3A) is highly expressed in SST;CR cells.
  - (G) Top: Schematic of GABA<sub>A</sub>R and ligand binding sites. Bottom: Differential expression of  $\alpha$ ,  $\beta$ , and  $\gamma$  subunits within a PCP; PVBCs and SST/CR cells have the most and least diversity, respectively.
  - (H) GABA<sub>A</sub>R subunit-level differences among PCPs. PVBCs have the highest levels of  $\alpha 1$ ,  $\alpha 4$ , and  $\alpha 5$  and also the inhibitory post-synaptic scaffolding protein Gphn (gephrin).
  - (I) Schematic comparison of neuromodulatory receptors among PVBCs, CCKCs, and LPCs.
  - (J) Differential expression of neuromodulatory receptors among PCPs; LPCs and CCKCs show the highest diversity.
  - (K) CGE-derived interneurons express more neuromodulatory receptors types than MGE-derived interneurons.
  - (L) Select neuromodulatory receptors specific to or enriched in LPCs.
  - (M) Differential expression of orphan GPCRs among PCPs shown as heatmap (left) and boxplots (right).
- For all boxplots, values on the y axis are in uTPM.  
See also Figure S4 and STAR Methods, Notched boxplots.



**Figure 5. Differential Expression of Regulatory Proteins in Second Messenger Pathways Customizes Intracellular Signaling in PCPs**

(A) A schematic showing that Ca<sup>2+</sup>, cAMP, cGMP, Ras, and Rho signaling pathways are differentially configured among PCPs. While the core skeletons of transduction machineries, kinase cascades, and effectors are common among PCPs (gray, low AUROC scores), a small set of regulatory proteins (red) are differentially expressed with high AUROC values.

(legend continued on next page)

that, while the G-protein subunits themselves are broadly expressed, regulators of G-protein signaling (RGS) family members manifest highly differential expression, often with binary on/off patterns among PCPs (AUROC = 0.93; **Figures 5A–5C**); this suggests that the turning off of  $G\alpha$  subunit, a crucial step of G-protein regulation, is implemented in a cell-type-specific manner. Downstream of G proteins, 7 of the 9 adenylate cyclase (ACs) members with different catalytic and regulatory properties are differentially expressed (AUROC = 0.85; **Figures 5A–5C**). More strikingly, phosphodiesterases (PDEs), which mediate rapid cAMP degradation (Maurice et al., 2014), is among the top differentially expressed gene families (AUROC = 0.94): 15 of the 22 members are differentially expressed, often with on/off patterns (**Figures 5A–5C** and S5A). The specific and correlated transcription of AC, PDE and their functional effectors in PCPs suggest possible mechanisms that craft the spatiotemporal patterns of a ubiquitous second messenger to direct cell and receptor (i.e., input) specific signal transduction, in part through subcellular targeted “signalosomes” containing particular members of synthetic and degradation enzymes with distinct catalytic and regulatory properties.

#### cGMP Signaling Modules in LPCs and CHCs

cGMP signaling in the brain is predominantly triggered by nitric oxide (NO). In mature cortex, the synthetic enzyme NOS1 is expressed at high levels in a small set of SST<sup>+</sup> long projection cells (LPCs) and much lower levels in several other GABA populations (Kilduff et al., 2011). We found that, in addition to NOS1, the neuronal L-arginine transporter Slc7a3 that supplies the substrate for NO synthesis (Friebe and Koesling, 2003) is also specific to LPC (**Figure 5E**); this tight co-expression may endow LPCs as the major source of cortical NO. As the key link from NO to cGMP production, the soluble guanylyl cyclase (sGC) functions as a strict heterodimer of  $\alpha$  and  $\beta$  subunits (Friebe and Koesling, 2003). While Gucy1 $\alpha$ 2 is expressed at low levels across PCPs, Gucy1 $\alpha$ 3 and Gucy1 $\beta$ 3 are highly enriched in CHCs, PVBCs, and LPCs but are nearly absent in SST/CR and VIP cells (**Figure 5E**). This suggests that cGMP signaling is prominent in the former three cell types but weak in the latter three populations. Consistently, cGMP-degrading Pde1a, 5a, and 11a are also highly enriched in LPCs and CHCs (**Figure 5E**), which may regulate the spatiotemporal dynamics of cGMP in these cells. Among the two types of cGMP-dependent PKGs, Prkg1 is found in all PCPs but with major enrichment in CHCs (**Figure 5E**). Furthermore, several PKG-regulated ion channels are also differentially enriched in these two cell types (**Figures 5F** and S5B). The stunning coordination in the expression of multiple (8–9) genes encoding almost the entire NO-cGMP pathway in LPCs and CHCs, from ligand synthesis and second

messenger signaling to potential effectors, suggests orchestration by a gene regulatory network.

#### Ras and Rho Small GTPases

Many cell surface receptors signal through a large set of Ras superfamily small GTPases to activate multiple kinase cascades that engage effectors, including transcription factors that regulate gene expression and cytoskeleton proteins that regulate cell shape, motility, adhesion, and intracellular transport (Alberts et al., 2014). The mammalian genome contains ~30 Ras-GTPases and ~20 Rho-GTPases, each is regulated by several dozens of guanine nucleotide exchange factors (GEFs) and inactivated by GTPase-activating proteins (GAPs) (Cherfils and Zeghouf, 2013). Within the Ras family, 21 of the 32 members showed major enrichment in specific PCPs (AUROC = 0.84). As different Ras family members might be activated by different upstream signals, have different cellular functions, and engage different downstream effectors (Alberts et al., 2014), PCPs might use Ras members to relay distinct external inputs and trigger appropriate transcription programs and other effectors that mediate long term cellular changes. Furthermore, both the Rho-GTPases and Rho-GEFs are differentially expressed. 37 of the 57 Rho-GEFs (AUROC = 0.82) and 14 of the 19 Rho-GTPases (AUROC = 0.72) are enriched in specific PCPs (**Figures 5D**, S5C, and S5D). As different Rho members are often activated by designated GEFs (Cook et al., 2014), differential expression of Rho signaling and regulatory components might provide the mechanism and capacity to maintain diversity of GABAergic neuron morphology, connectivity, and to support different forms of neurite and synaptic motility and plasticity.

Together, our results suggest that, among the vast number of intracellular signaling proteins constituting myriad pathways that transduce major categories of extracellular inputs, a small number of regulatory components encoded by just a few gene families are differentially expressed among PCPs and likely customize a handful of second messenger pathways. Superimposed upon the core common signaling scheme, these regulatory components likely shape the specificity and spatiotemporal dynamics of broadly acting second messengers to translate specific inputs to appropriate cellular responses.

#### Differential Expression of Neuropeptides and Vesicle Release Machinery Shape Distinct Outputs

The single most important physiological action of a nerve cell is influencing the activity of its target cells through the release of appropriate neurochemicals in appropriate “styles.” We discovered a surprising diversity of neurochemical contents among PCPs and correlated differential expression of components of

(B) A GPCR signaling module illustrating that while multiple components (gray) are common among PCPs, different members of key regulatory proteins, such as RGS, AC, PDE, and AKAPs, are differentially expressed.

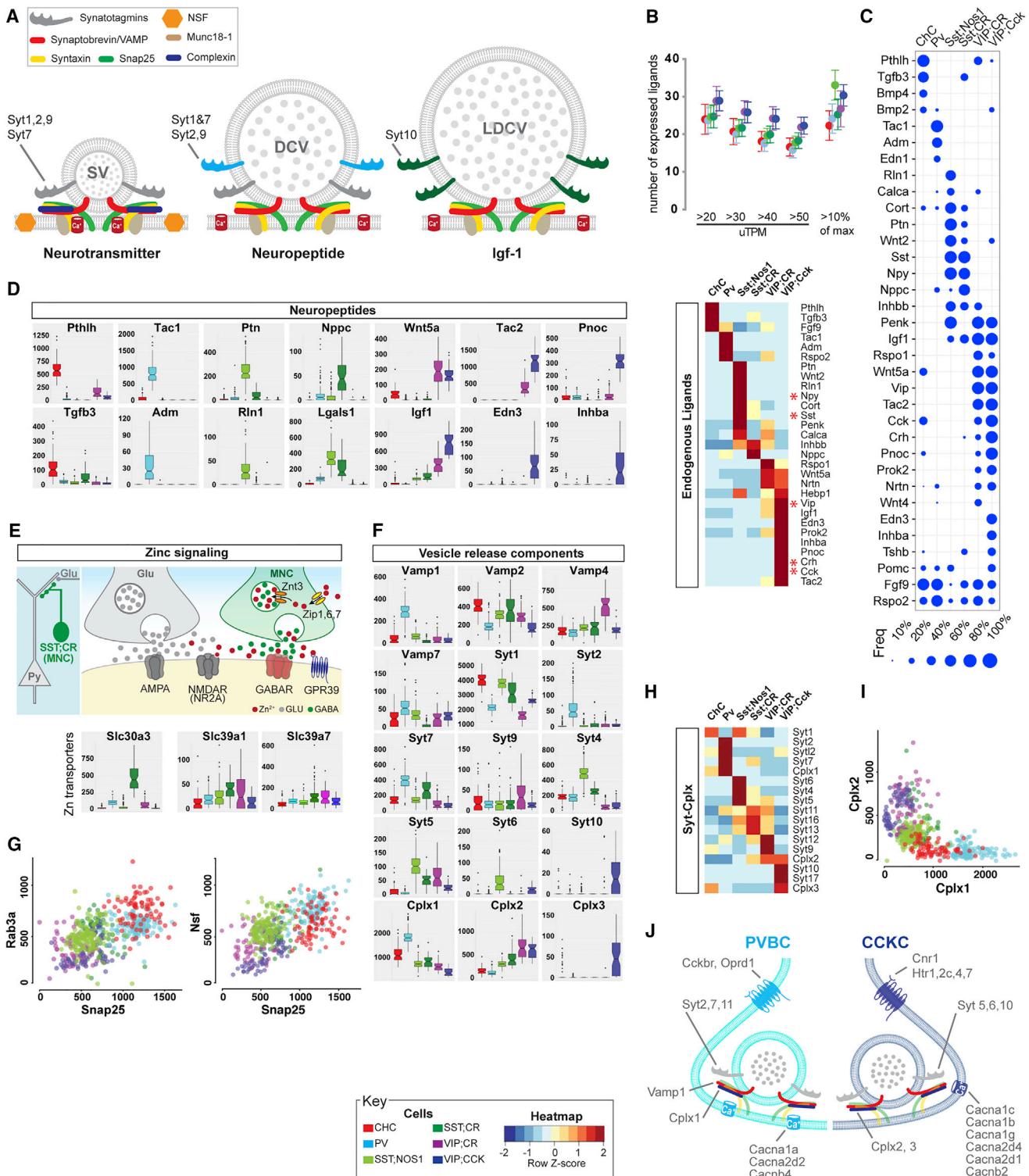
(C) Different combinations of RGS, AC, and PDE members are enriched in individual PCPs.

(D) Differential expression of several classes of signaling proteins with high AUROC scores.

(E) Predicted NO-cGMP signaling in SST;NOS1 and CHC cells. The entire pathway of NO synthesis and cGMP production (guanylyl cyclase), degradation (PDE), kinase signaling (PKG), and putative phosphorylation targets are coherently expressed or enriched in SST;NOS1 cells and CHCs.

(F) Differential expression of key components of NO-cGMP signaling (depicted in E) among PCPs; note on/off patterns or dramatic level differences. For all boxplots, values on the y axis are in uTPM.

See also **Figure S5** and **STAR Methods**, Notched boxplots.



**Figure 6. Differential Expression of Neuropeptides and Vesicle Release Machinery Shape Outputs and Release Styles in PCPs**

(A) Schematic of vesicular release machinery for synaptic vesicle (neurotransmitters), dense core vesicle (neuropeptides), and large dense core vesicle (protein/hormones) with putative Syt members.

(B) Top: Each PCP is estimated to express 20–30 peptides based on either a sliding threshold or dynamic threshold (10% of max expression value). Bottom: Differential expression of endogenous ligands that constitute a neuropeptide code for PCPs. Error bars show SD.

(legend continued on next page)

vesicular release machinery that may contribute to different release styles (Figure 6A).

#### A Neuropeptide Code of PCPs

The release of different transmitters, peptides, and hormones represent a fundamental distinction among neuron types as they produce categorically different outputs that elicit distinct physiological actions in target cells. We revealed a neuropeptide code of GABAergic neurons. Over 30 neuropeptides, hormones, and secreted ligands are expressed in over 50% of single cells of the PCPs, and each expresses ~3–10 different endogenous ligands (Figures 6B and S6A); individual neurons express multiple peptide and protein ligands (Figure 6C). Importantly, differential expression of these ligands is the most discriminating gene family for PCPs (AUROC = 0.96). Multiple PCPs are uniquely marked by individual ligands (Figures 6B and 6C) (e.g., CHC: Pthlh, PV: Tac1, Adm, NOS1/SST: Ptn, Rln1, CR/SST: Nppc, VIP/CCK: Edn3, Pnoc). These results indicate that, beyond their morphophysiological differences, PCPs are different neuroendocrine cells that produce distinct chemical outputs and elicit distinct physiological effects. Consistent with the demand for processing and packaging diverse neuropeptides, the granin gene family, which regulates pre-prohormone cleavage and biogenesis of DCVs, also shows differential expression (AUROC = 0.81; Table S4). Further, based on the role pleiotropin (PTN) in axon myelination and its exclusive expression in NOS1/SST cells, we discovered that the axons of these long projection GABAergic neurons are myelinated (Figures S6B–S6G).

#### Vesicular Zinc Transporter in SST/CR Cells Suggests Co-release of GABA and Zinc

The divalent cation zinc is enriched in cerebral hemisphere and acts as a potent modulator of neuronal signaling (Marger et al., 2014). The vesicular transporter ZnT3 in certain glutamatergic neurons enables zinc co-release with glutamate. Synaptic source of zinc modulates multiple ion channels (e.g., inhibiting extra-synaptic NMDA receptors) (Marger et al., 2014). Surprisingly, we discovered that ZnT3 (Slc30a3) is highly and specifically expressed in SST/CR cells (Figure 6E), along with Zip1 (Slc39a1) and Zip7a (Slc39a7) transporters that uptake extracellular zinc into the cytosol. These results suggest that SST/CR cells co-release zinc and GABA. As most SST/CR cells are Martinotti cells that target the distal dendrites and spines of pyramidal neurons with abundant NMDARs (Silberberg and Markram, 2007), their powerful dendritic inhibition might be mediated through two parallel mechanisms: synaptic activation of GABA<sub>A</sub>Rs by GABA and extra-synaptic inhibition of NMDARs by zinc.

#### Syaptotagmin Members Correlate with Vesicular Contents

The molecular components of vesicle fusion machinery are encoded by several multi-gene families (Südhof, 2013), but it is not clear how different gene family members shape transmitter and neuropeptide release properties (Moghadam and Jackson, 2013). We revealed comprehensive molecular profiles of vesicle release machinery in PCPs (Figures 6F–6I). Several core components of the fusion complex and active zone are broadly expressed, including syntaxins (AUROC = 0.5), SNAP complex (AUROC = 0.614), RIMs and RIM binding proteins (AUROC = 0.57) (Table S4). Yet even among these core components, VAMP (synaptobrevins) and SNAP members are significantly enriched in specific PCPs (Figures 6F–6H). The vesicular Ca<sup>2+</sup> sensor synaptotagmins (Syt) show more distinct patterns: 14 of the 17 Syts are differentially expressed among PCPs (Figures 6G–6I) (AUROC = 0.78); individual neurons express 6–9 Syts. In particular, VIP/CCKCs specifically express Syt10 that mediates the release of Igf1 (Cao et al., 2011), which is also highly enriched in the same cells (Figures 6F and 6D). These results suggest that each PCP might deploy a specific set of Syts with different sensitivity to spatiotemporal Ca<sup>2+</sup> signals that trigger particular types of fusion reactions, thereby shaping the specificity and properties in parallel exocytosis pathways.

#### Molecular Signatures of Vesicular Release Styles

The release sites (e.g., synaptic versus non-synaptic), temporal characteristics (e.g., fast-synchronous versus slow-asynchronous), and short-term dynamics (facilitating versus depressing) of GABA and neuropeptides produce distinct spatiotemporal patterns of receptor activation and post-synaptic cell firing that impact circuit computation (Armstrong and Soltesz, 2012; Markram et al., 2015). Our analyses begin to reveal molecular distinctions of fast-synchronous (in PVBCs) versus slow-sustained release (in CCKBCs) machinery, which manifests even at the level of core fusion complex. Vamp1, Snap25, Nsf, and Rab3a are highly enriched in PV compared to VIP/CCKCs (Figures 6G and 6H); co-expression of all three components in PVBCs supports their fast release properties (Parpura and Mohideen, 2008). Among the Ca<sup>2+</sup>-binding Syts localized to SVs (Syt1, Syt2, and Syt9), both PV and VIP/CCKCs (and other PCPs) express Syt1, but only PVBCs express Syt2, which exhibits the fastest onset and decline in release. PVBCs express highest level of Cplx1 but lowest level of Cplx2, while VIP/CCKCs show the opposite (Figures 6H–6J). This Cplx profile is highly congruent with the Syt profile, as Cplx1 is implicated in fast and synchronous release and in clamping spontaneous release

(C) Fraction of individual cells expressing the most common neuropeptides among PCPs. Dot size represents fraction (see key).

(D) On/off expression (uTPM) of specific neuropeptides/endogenous ligands; the gene family that best distinguishes PCPs (AUROC = 0.96).

(E) Schematic showing that zinc may be co-released with GABA from SST;CR terminals. While GABA acts on GABA<sub>A</sub>Rs, zinc may act on nearby non-synaptic NMDARs and influence glutamatergic transmission. Boxplot: high-level specific expression of the zinc vesicular transporter Slc30a3 in SST;CR cells, which also contain the zinc uptake importers Slc39a1 and Slc39a7.

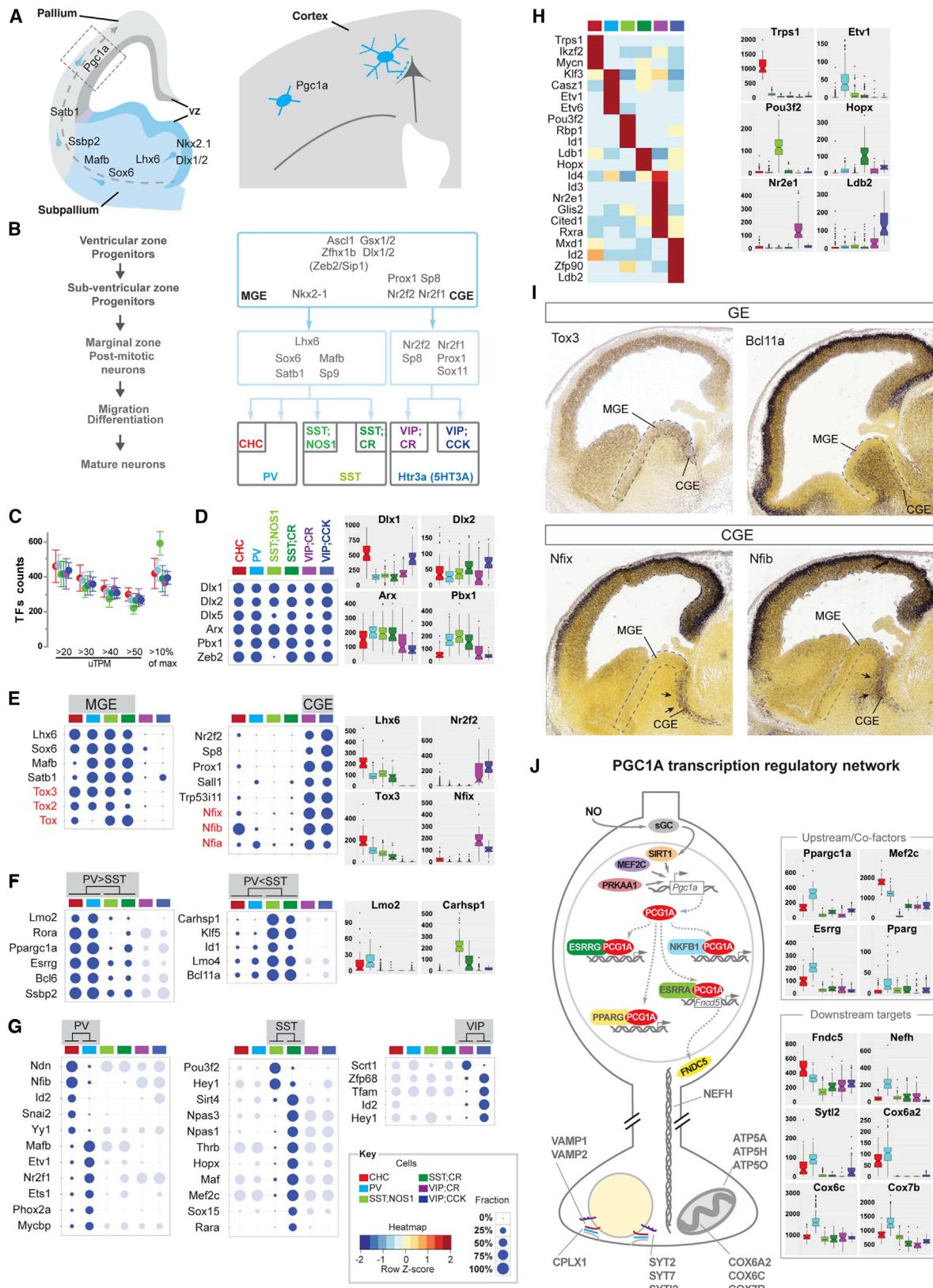
(F) Differential expression of vesicle release machinery components suggest different release styles in Ca<sup>2+</sup> sensitivity and dynamics among PCPs.

(G) Scatterplots of mRNA levels (uTPMs) of Snap25 versus Rab3a (left) and Snap25 versus Nsf (right).

(H) Selective expression of synaptotagmin and complexin families in PCPs.

(I) Scatterplot of Cplx1 versus Cplx2 levels shows that fast-release synapses of PV and CHCs are biased toward Cplx1, whereas slow-release synapses of CCKCs mainly utilize Cplx2.

(J) Molecular correlates of fast-synchronous and slow-sustained vesicle release mechanisms in PVBCs and CCKCs with contrasting GABA release styles. See also Figure S6 and STAR Methods, Notched boxplots.



(legend on next page)

(Yang et al., 2013). These results suggest that fast-synchronous release is supported by high levels of VAMP1, SNAP25, NSF, SYT2, and CPLX1, whereas slow- asynchronous release is shaped with low levels of these components and high levels of CPLX2. Together with co-expression of matching properties of  $\text{Ca}^{2+}$  channels and CaBPs, these results suggest that PCPs might transmit multiple neurochemicals in multiple release styles through differential and coordinated expression of gene families that customize the vesicle fusion machinery (Table S7).

In addition, Syt4, Syt5, and Syt6 are highly or uniquely enriched in NOS1/SST long projection neurons, which express over 11 peptides (Figure 6B–6D and 6F). It is possible that these uncharacterized Syts might mediate synaptic release of peptide-containing DVCs or their endocrine/paracrine release along the axon-dendritic membrane.

### Transcription Factors Register the Developmental History and Contribute to the Maintenance of PCP Phenotypes

#### GABAergic Neurons Retain a Transcription Resume that Registers Their Developmental History

Previous studies suggest that neuronal identities can be maintained by sustained expression of the same set of transcription factors that initiate terminal differentiation during development (Dalla Torre di Sanguinetto et al., 2008; Deneris and Hobert, 2014). In embryonic subpallium, transcriptional cascades orchestrate the specification and differentiation of major clades of GABAergic neurons (Kepcs and Fishell, 2014; Nord et al., 2015) (Figures 7A and 7B). We found that each PCP expresses ~350–400 TFs and over 300 TFs are expressed in an individual cell (Figure 7C). Among ~34 TF classes, basic-helix-loop-helix proteins, nuclear hormone receptors, POU-homeoboxes, and kruppel-like transcription factors are most differentially expressed among PCPs (Tables S3 and S4), and multiple TFs individually marks each PCP (Figures 7H and S7B).

As the 6 PCPs are embedded in 3 non-overlapping populations (PV, SST, VIP) derived from 2 separate developmental origins (MGE versus CGE) (Figure 7B), this data structure establishes a link between TF profiles in mature neurons to those in their embryonic precursors with cell-type and single-cell resolution. Almost all well-studied TFs in embryonic precursors main-

tain expression within the same clade of mature PCPs (Figures 7D and 7E): whereas Lhx6, Sox6, Mafb, and Satb1 are expressed in PV and SST populations (the MGE clade), Coup-TF2, Sp8, Prox1, Npas1, and Npas3 are expressed in VIP populations (the CGE clade). We further discovered additional TFs with similar patterns: whereas Tox family members (Tox, Tox2, and Tox3) (Artegiani et al., 2015) are restricted to the MGE clade, Nfi family members (Nfia', Nfib', and Nfix) (Piper et al., 2014), Sall1 and Trp53i11 are restricted to the CGE clade. Importantly, by “reverse tracking” of their developmental history through screening the Allen Developmental Mouse Brain Atlas, we found that each of these TFs is indeed expressed in the embryonic MGE or CGE, consistent with their clade relationship (Figures 7I and S7C).

Furthermore, by hierarchical and pairwise comparison, we defined multiple sets of TFs that distinguish PV versus SST population (Figure 7F), and the PV, SST, and VIP pairs of PCPs (Figure 7G). Again, we found evidence for developmental continuity of TF expression from embryonic precursors to mature neurons (Figure S7C). In particular, multiple PVBC-enriched TFs initiate expression at different stages of development, from Lhx6 and its downstream cascade (Figures 7A, 7D, and 7E) to Ssbp2, Nfib, and PGC1 $\alpha$  (Figures 7E–7G; (Batista-Brito et al., 2008), indicating that they maintain sequentially acquired transcriptional factors (Figure 7A).

These results suggest that PCPs register the developmental history of their transcription program, i.e., a “transcription resume,” through sustained expression of sequentially accumulated transcription factors. Conditional deletions of several of these TFs in adult cortex result in the loss of markers and physiological properties characteristic to their identity (e.g., (Close et al., 2012; Touzot et al., 2016), suggesting their requirement in mature neurons to maintain aspects of cell phenotypes and identity.

#### The PGC1 $\alpha$ Transcription Program Coordinates the Release and Metabolic Properties of PV Cells

The fast input-output transformation of PVBCs requires specialized energy metabolism and mitochondria features (Lucas et al., 2014). The transcription coactivator PGC1 $\alpha$  cooperates with multiple transcription factors to regulate mitochondria biogenesis and energy metabolism (Lin et al., 2005). In the brain, it is

**Figure 7. Transcription Factor Profiles Register the Developmental History and Contribute to Maintenance of PCP Phenotypes**

- (A) Schematic developmental trajectory of cortical GABAergic neurons (PVBCs, for example), with TFs expressed at different stages.
  - (B) Schematic of MGE and CGE transcription cascades that regulate the development of different clades of GABAergic neurons, including PCPs.
  - (C) Each PCP is estimated to express ~400 TFs. Error bars show SD.
  - (D–G) Fraction of cells expressing a given TF (10% of max level); boxplots show expression levels of selected TFs. (D) TFs in subpallium progenitors and GABA neuron precursors maintain their expression in PCPs in adult. (E) TFs expressed in early post-mitotic MGE- and CGE-derived neurons maintain expression within same clade of PCPs. Embryonic expression of Tox and Nfi family TFs were deduced from transcriptome analysis of PCPs and confirmed in (I). (F) Among MGE-derived PCPs, subsets of TFs are preferentially expressed in PV (PV > SST) or SST (PV < SST) groups; CGE groups are not compared and shown in light shade; selected boxplots are shown. (G) Within the PV, SST, and VIP group, subsets of TFs are enriched in one or the other PCP; PCPs that are not compared are shown in light shade.
  - (H) Differential expression of TFs is largely exclusive to each PCP (left); examples of on/off expression in individual PCPs (right).
  - (I) Retrospective screen of Allen Developmental Mouse Brain *in situ* database reveals that several TFs that express in MGE- or CGE-derived PCPs identified by transcriptome analysis indeed begin their expression in the corresponding embryonic germinal zone.
  - (J) Schematic of the Ppargc1 $\alpha$  (PGC1 $\alpha$ ) transcription regulatory network highly enhanced in PVBCs. Multiple PGC1 $\alpha$  upstream TF activators, cofactors, and large fraction (>75%) of downstream effectors are enriched over 1.5-fold in PVBCs ( $p < 5.0^{-7}$ ). Boxplots show different expression levels of select sets of PGC1 $\alpha$  co-factors and targets and putative targets among PCPs.
- See also Figure S7 and STAR Methods, Notched boxplots.

largely restricted to cortical PVBCs and may directly regulate genes involved in mitochondria function and transmitter release (Lucas et al., 2014). Several PGC1a co-factors (e.g., Rora, Esrrg, and Pparg) and all of its potential targets (PV, Syt2, Nefh, Cplx1, Atp50, and Atp5a1) (Lucas et al., 2014) are substantially enriched in PVBCs (Figures 7J and S7D). We further found an extended set of PVBC enriched mRNAs associated with metabolic and mitochondria pathways (e.g., Fndc5, Cox6a2, Cox6c, and Cox7b) (Figure 7J). These results suggest that PGC1 $\alpha$  might organize a transcription module that coordinates the release and metabolic properties in PVBCs.

### Molecular Portraits of GABAergic Cell Types

We have discovered highly correlated and congruent gene expression across multiple gene families and functional categories in each PCP. While coordinated expression within a functional category may shape a specific cell property, those across categories may jointly shape a set of congruent cell properties that together characterize cell phenotypes and identity (Table S7).

## DISCUSSION

### Neuronal Cell Types Are Defined by Transcriptional Signatures of Input-Output Communication

The biological basis and mechanistic framework of neuronal identity and diversity have been elusive. A communication-based approach to defining neuron types operationally by input-output relationships has been proposed (Lerner et al., 2016). Here, through single-cell transcriptome analysis of phenotype-characterized GABAergic neurons, we have discovered that transcriptional architecture of input-output (I/O) synaptic communication may underlie neuronal identity. This overarching and mechanistic definition of neuronal identity integrates cell phenotypes along multiple axes and provides a foundation for understanding neuronal diversity and achieving biological (i.e., beyond operational) classification.

Although morphology is a common and intuitive description of neurons, it reflects and serves the more fundamental purpose of achieving proper connectivity. Thus morphological variability likely belies the co-variation of pre- and post-synaptic neurites that preserves connectivity patterns (Seung and Sümbül, 2014). Indeed, morphological types can be reliably identified from dense connectomes by computational algorithms (Jonas and Kording, 2015). Beyond anatomical connectivity, the physiological operation of a neuron type transforms information contents embedded in its synaptic inputs (e.g., transmitter and modulator types, strength, and spatiotemporal dynamics) to appropriate outputs (Kepecs and Fishell, 2014), which are often characterized by cell intrinsic style of neurochemical release (e.g., vesicle contents and release speed, dynamics, and plasticity). Although highly valuable, most electrophysiological measurements at cell soma regions, often in artificial conditions, provide a limited window into the elaborate subcellular biophysical, signaling, and metabolic processes. Our comprehensive transcription overview of the synaptic, intrinsic, and release machineries reveals strikingly coherent molecular ensemble properties congruent with well-characterized physiological, bio-

physical, and release properties of PCPs. They further predict multiple novel physiological features that can be experimentally verified. Thus transcriptional signatures of synaptic I/O machineries may begin to harmonize and extend the hitherto often limited, disparate, and technically challenging electrophysiological measurements. Furthermore, neuron types defined by connectivity pattern and I/O styles may represent distinct structural and physiological motifs, with characteristic sets of dynamic properties that support and constrain their roles in circuit operations. Task-dependent recruitment of these motifs into brain networks may engage their systems level information processing and function. Finally, although transcription is influenced by cellular milieu including neural activity, core features of transcriptomes are outputs of cellular epigenomes customized primarily through developmental programming of the genome. Therefore, transcriptional signatures of synaptic I/O communication may integrate anatomical, physiological, functional, and developmental genetic features that together define neuron types.

### Computation Genomics Screen of Gene Ensembles that Contribute to Cell Phenotypes

Previous studies often identify molecular markers of “transcriptomic types” that do not readily inform cell phenotypes. We instead focused on analyzing gene ensemble profiles (i.e., gene families) encoding proteins that constitute cellular modules (Hartwell et al., 1999) (i.e., macromolecular machines, signaling complexes), which more readily explain and predict cell phenotypes. Leveraging substantial phenotypic information of PCPs as an assay, MetaNeighbor allowed us to systematically screen all gene families and rank their ability to discriminate PCPs. This enabled us to discover a rather small set of functionally related gene families, which likely shape the I/O communication patterns of PCPs.

### Transcription Resume May Reflect the Developmental Accumulation of Gene Regulatory Programs that Maintain Cell Phenotypes

Transcriptional control of “gene batteries” has been shown to coordinate effector gene expression and cellular properties (Hobert et al., 2010). Our study substantiates the role PGC1 $\alpha$  transcription program in coordinating multiple physiological and metabolic properties in PVBCs (Figure 7J). Together with results of highly congruent expression of functional gene ensembles in other PCPs (e.g., NO-cGMP signaling pathway in LPCs), the cumulative evidence precludes piece meal, coincidental mechanisms of co-expression and suggests transcriptional programs that maintain cell phenotypes and identity.

We found that numerous developmental transcription programs initiated at successive stages of post-mitotic differentiation are maintained in the same clade of mature neurons (Figures 7A–7G). It is possible that a hierarchy of transcription programs progressively restricts cell fate and establishes their competence in responding to subsequently initiated programs that guide the migration, differentiation, and maturation of GABAergic neurons. Our results further suggest that most developmental programs persist in mature neurons and constitute a transcriptional resume that may maintain cell phenotypes and identity.

CHCs and PVBCs are implicated in the pathophysiology of schizophrenia (Gonzalez-Burgos et al., 2015), and mental disorders likely result from altered neural connectivity rooted in deficient cell types. Key transcription profiles of increasing number of neuron types will facilitate identifying homologous cell types across species through conserved genetic features, linking altered gene expression to aberrant cellular and circuit properties, and discovering therapeutic targets to ameliorate circuit deficits.

## STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Animals
- METHOD DETAILS
  - Phenotype characterized GABAergic subpopulations
  - RNA double *in situ* and imaging
  - Super-resolution microscopy
  - Manual cell sorting
  - Linear RNA amplification
  - cDNA library generation and sequencing
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Mapping and tag counting
  - Fisher's meta-analytic Differential Expression
  - MetaNeighbor
  - Notched Boxplots
- DATA AND SOFTWARE AVAILABILITY
  - Ethics

## SUPPLEMENTAL INFORMATION

Supplemental Information includes seven figures and seven tables and can be found with this article online at <http://dx.doi.org/10.1016/j.cell.2017.08.032>.

## AUTHOR CONTRIBUTIONS

Z.J.H. and A.P. conceived the study. Z.J.H. organized the study. A.P. performed scRNA-seq experiments. A.P. and R.R. performed mRNA *in situ* and histology experiments. A.P., Z.J.H., J.G., and M.C. analyzed the data. M.C. and J.G. developed MetaNeighbor and performed differential expression and AUROC analysis. M.H. contributed to driver lines. Z.J.H. and A.P. wrote the manuscript with contributions from J.G. and M.C.

## ACKNOWLEDGMENTS

We thank M. Wigler for UMs, J. Kendall for coding, N. El-Amine for microscopy; and L. Luo, C. McBain, A. Zador, L. Van Aelst, and G. Fishell for comments, and A. Kepcs and M. Rosbash for advice. This work was supported by grants from the NIH (5R01MH094705-04 and R01MH109665-01 to Z.J.H.; 1R01MH113005 to J.G.; and 1F32MH114501 to M.C.), by the CSHL Robertson Neuroscience Fund (to Z.J.H.), and by a NARSAD Post-doctoral Fellowship (to A.P.).

Received: December 21, 2016

Revised: May 4, 2017

Accepted: August 16, 2017

Published: September 21, 2017

## REFERENCES

- Akgül, G., and McBain, C.J. (2016). Diverse roles for ionotropic glutamate receptors on inhibitory interneurons in developing and adult brain. *J. Physiol.* 594, 5471–5490.
- Alberts, B., Johnson, A., Lewis, J., Morgan, D., Raff, M., Roberts, K., and Walter, P. (2014). Cell signaling. In *Molecular Biology of the Cell*, B. Alberts, ed. (Garland Science), pp. 850–866.
- Armañanzas, R., and Ascoli, G.A. (2015). Towards the automatic classification of neurons. *Trends Neurosci.* 38, 307–318.
- Armstrong, C., and Soltesz, I. (2012). Basket cell dichotomy in microcircuit function. *J. Physiol.* 590, 683–694.
- Artegiani, B., de Jesus Domingues, A.M., Bragado Alonso, S., Brandl, E., Massalini, S., Dahl, A., and Calegari, F. (2015). Tox: a multifunctional transcription factor and novel regulator of mammalian corticogenesis. *EMBO J.* 34, 896–910.
- Batista-Brito, R., Machold, R., Klein, C., and Fishell, G. (2008). Gene expression in cortical interneuron precursors is prescient of their mature function. *Cereb. Cortex* 18, 2306–2317.
- Cajal, S.R.y. (1892). El nuevo concepto de la histología de los centros nerviosos. *Rev. Cienc. Med.* 18, 457–476.
- Cao, P., Maximov, A., and Südhof, T.C. (2011). Activity-dependent IGF-1 exocytosis is controlled by the Ca(2+)-sensor synaptotagmin-10. *Cell* 145, 300–311.
- Chambers, John M. (1983). Graphical methods for data analysis (Wadsworth International Group), p. 4, 47.
- Cherfils, J., and Zeghouf, M. (2013). Regulation of small GTPases by GEFs, GAPs, and GDIs. *Physiol. Rev.* 93, 269–309.
- Chorin, E., Vinograd, O., Fleidervish, I., Gilad, D., Herrmann, S., Sekler, I., Aizenman, E., and Hershkoff, M. (2011). Upregulation of KCC2 activity by zinc-mediated neurotransmission via the mZnR/GPR39 receptor. *J. Neurosci.* 31, 12916–12926.
- Close, J., Xu, H., De Marco García, N., Batista-Brito, R., Rossignol, E., Rudy, B., and Fishell, G. (2012). Satb1 is an activity-modulated transcription factor required for the terminal differentiation and connectivity of medial ganglionic eminence-derived cortical interneurons. *J. Neurosci.* 32, 17690–17705.
- Cook, D.R., Rossman, K.L., and Der, C.J. (2014). Rho guanine nucleotide exchange factors: regulators of Rho GTPase activity in development and disease. *Oncogene* 33, 4021–4035.
- Crow, M., Paul, A., Ballouz, S., Huang, Z.J., and Gillis, J. (2016). Exploiting single-cell expression to characterize co-expression replicability. *Genome Biol.* 17, 101.
- Crow, M., Paul, A., Ballouz, S., Huang, Z.J., and Gillis, J. (2017). Addressing the looming identity crisis in single cell RNA-seq. *bioRxiv*. <http://dx.doi.org/10.1101/150524>.
- Dalla Torre di Sanguinetto, S.A., Dasen, J.S., and Arber, S. (2008). Transcriptional mechanisms controlling motor neuron diversity and connectivity. *Curr. Opin. Neurobiol.* 18, 36–43.
- de Wit, J., and Ghosh, A. (2016). Specification of synaptic connectivity by cell surface interactions. *Nat. Rev. Neurosci.* 17, 22–35.
- DeFelipe, J., López-Cruz, P.L., Benavides-Piccione, R., Bielza, C., Larrañaga, P., Anderson, S., Burkhalter, A., Cauli, B., Fairén, A., Feldmeyer, D., et al. (2013). New insights into the classification and nomenclature of cortical GABAergic interneurons. *Nat. Rev. Neurosci.* 14, 202–216.
- Deneris, E.S., and Hobert, O. (2014). Maintenance of postmitotic neuronal cell identity. *Nat. Neurosci.* 17, 899–907.
- Dixon, C., Sah, P., Lynch, J.W., and Keramidas, A. (2014). GABAA receptor  $\alpha$  and  $\gamma$  subunits shape synaptic currents via different mechanisms. *J. Biol. Chem.* 289, 5399–5411.
- Eberwine, J., Yeh, H., Miyashiro, K., Cao, Y., Nair, S., Finnell, R., Zettel, M., and Coleman, P. (1992). Analysis of gene expression in single live neurons. *Proc. Natl. Acad. Sci. USA* 89, 3010–3014.

- Fame, R.M., Dehay, C., Kennedy, H., and Macklis, J.D. (2017). Subtype-specific genes that characterize subpopulations of callosal projection neurons in mouse identify molecularly homologous populations in macaque cortex. *Cereb Cortex* 27, 1817–1830.
- Friebe, A., and Koesling, D. (2003). Regulation of nitric oxide-sensitive guanylyl cyclase. *Circ. Res.* 93, 96–105.
- Gonzalez-Burgos, G., Cho, R.Y., and Lewis, D.A. (2015). Alterations in cortical network oscillations and parvalbumin neurons in schizophrenia. *Biol. Psychiatry* 77, 1031–1040.
- Haering, S.C., Tapken, D., Pahl, S., and Hollmann, M. (2014). Auxiliary subunits: shepherding AMPA receptors to the plasma membrane. *Membranes (Basel)* 4, 469–490.
- Halls, M.L., and Cooper, D.M. (2011). Regulation by Ca<sup>2+</sup>-signaling pathways of adenylyl cyclases. *Cold Spring Harb. Perspect. Biol.* 3, a004143.
- Hartwell, L.H., Hopfield, J.J., Leibler, S., and Murray, A.W. (1999). From molecular to modular cell biology. *Nature* 402 (6761, Suppl), C47–C52.
- Hashimshony, T., Wagner, F., Sher, N., and Yanai, I. (2012). CEL-Seq: single-cell RNA-seq by multiplexed linear amplification. *Cell Rep.* 2, 666–673.
- He, M., Tucciarone, J., Lee, S., Nigro, M.J., Kim, Y., Levine, J.M., Kelly, S.M., Krugikov, I., Wu, P., Chen, Y., et al. (2016). Strategies and tools for combinatorial targeting of GABAergic neurons in mouse cerebral cortex. *Neuron* 91, 1228–1243.
- Hobert, O., Carrera, I., and Stefanakis, N. (2010). The molecular and gene regulatory signature of a neuron. *Trends Neurosci.* 33, 435–445.
- Hu, H., Gan, J., and Jonas, P. (2014). Interneurons. Fast-spiking, parvalbumin<sup>+</sup> GABAergic interneurons: from cellular design to microcircuit function. *Science* 345, 1255263.
- Huang, Z.J., and Zeng, H. (2013). Genetic approaches to neural circuits in the mouse. *Annu. Rev. Neurosci.* 36, 183–215.
- Jonas, E., and Kording, K. (2015). Automatic discovery of cell types and microcircuitry from neural connectomics. *eLife* 4, e04250.
- Kepcs, A., and Fishell, G. (2014). Interneuron cell types are fit to function. *Nature* 505, 318–326.
- Kerti-Szigeti, K., Nusser, Z., and Eyre, M.D. (2014). Synaptic GABA<sub>A</sub> receptor clustering without the γ2 subunit. *J. Neurosci.* 34, 10219–10233.
- Kilduff, T.S., Cauli, B., and Gerashchenko, D. (2011). Activation of cortical interneurons during sleep: an anatomical link to homeostatic sleep regulation? *Trends Neurosci.* 34, 10–19.
- Kolodkin, A.L., and Tessier-Lavigne, M. (2011). Mechanisms and molecules of neuronal wiring: a primer. *Cold Spring Harb. Perspect. Biol.* 3, 3.
- Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9, 357–359.
- Lerner, T.N., Ye, L., and Deisseroth, K. (2016). Communication in neural circuits: tools, opportunities, and challenges. *Cell* 164, 1136–1150.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* 25, 2078–2079.
- Lin, J., Handschin, C., and Spiegelman, B.M. (2005). Metabolic control through the PGC-1 family of transcription coactivators. *Cell Metab.* 1, 361–370.
- Lucas, E.K., Dougherty, S.E., McMeekin, L.J., Reid, C.S., Dobrunz, L.E., West, A.B., Hablitz, J.J., and Cowell, R.M. (2014). PGC-1α provides a transcriptional framework for synchronous neurotransmitter release from parvalbumin-positive interneurons. *J. Neurosci.* 34, 14375–14387.
- Marger, L., Schubert, C.R., and Bertrand, D. (2014). Zinc: an underappreciated modulatory factor of brain function. *Biochem. Pharmacol.* 91, 426–435.
- Markram, H., Muller, E., Ramaswamy, S., Reimann, M.W., Abdellah, M., Sanchez, C.A., Allamaki, A., Alonso-Nanclares, L., Antille, N., Arsever, S., et al. (2015). Reconstruction and simulation of neocortical microcircuitry. *Cell* 163, 456–492.
- Maurice, D.H., Ke, H., Ahmad, F., Wang, Y., Chung, J., and Manganiello, V.C. (2014). Advances in targeting cyclic nucleotide phosphodiesterases. *Nat. Rev. Drug Discov.* 13, 290–314.
- Moghadam, P.K., and Jackson, M.B. (2013). The functional significance of synaptotagmin diversity in neuroendocrine secretion. *Front. Endocrinol. (Lausanne)* 4, 124.
- Nord, A.S., Pattabiraman, K., Visel, A., and Rubenstein, J.L. (2015). Genomic perspectives of transcriptional regulation in forebrain development. *Neuron* 85, 27–47.
- Olsen, R.W., and Sieghart, W. (2008). International Union of Pharmacology. LXX. Subtypes of gamma-aminobutyric acid(A) receptors: classification on the basis of subunit composition, pharmacology, and function. *Update. Pharmacol. Rev.* 60, 243–260.
- Parpura, V., and Mohideen, U. (2008). Molecular form follows function: (un)snaring the SNAREs. *Trends Neurosci.* 31, 435–443.
- Piper, M., Barry, G., Harvey, T.J., McLeay, R., Smith, A.G., Harris, L., Mason, S., Stringer, B.W., Day, B.W., Wray, N.R., et al. (2014). NFIB-mediated repression of the epigenetic factor Ezh2 regulates cortical development. *J. Neurosci.* 34, 2921–2930.
- Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., et al. (2012). Fiji: an open-source platform for biological-image analysis. *Nat Methods* 9, 676–682.
- Seung, H.S., and Sümbül, U. (2014). Neuronal cell types and connectivity: lessons from the retina. *Neuron* 83, 1262–1272.
- Shekhar, K., Lapan, S.W., Whitney, I.E., Tran, N.M., Macosko, E.Z., Kowalczyk, M., Adiconis, X., Levin, J.Z., Nemesh, J., Goldman, M., et al. (2016). Comprehensive classification of retinal bipolar neurons by single-cell transcriptomics. *Cell* 166, 1308–1323.e1330.
- Silberberg, G., and Markram, H. (2007). Disynaptic inhibition between neocortical pyramidal cells mediated by Martinotti cells. *Neuron* 53, 735–746.
- Staiger, J.F., Masanneck, C., Schleicher, A., and Zuschratter, W. (2004). Calbindin-containing interneurons are a target for VIP-immunoreactive synapses in rat primary somatosensory cortex. *J. Comp. Neurol.* 468, 179–189.
- Südhof, T.C. (2013). Neurotransmitter release: the last millisecond in the life of a synaptic vesicle. *Neuron* 80, 675–690.
- Sugino, K., Hempel, C.M., Miller, M.N., Hattox, A.M., Shapiro, P., Wu, C., Huang, Z.J., and Nelson, S.B. (2006). Molecular taxonomy of major neuronal classes in the adult mouse forebrain. *Nat. Neurosci.* 9, 99–107.
- Takahashi, H., and Craig, A.M. (2013). Protein tyrosine phosphatases PTPδ, PTPσ, and LAR: presynaptic hubs for synapse organization. *Trends Neurosci.* 36, 522–534.
- Taniguchi, H., Lu, J., and Huang, Z.J. (2013). The spatial and temporal origin of chandelier cells in mouse neocortex. *Science* 339, 70–74.
- Tasic, B., Menon, V., Nguyen, T.N., Kim, T.K., Jarsky, T., Yao, Z., Levi, B., Gray, L.T., Sorensen, S.A., Dolbeare, T., et al. (2016). Adult mouse cortical cell taxonomy revealed by single cell transcriptomics. *Nat. Neurosci.* 19, 335–346.
- Touzot, A., Ruiz-Reig, N., Vitalis, T., and Studer, M. (2016). Molecular control of two novel migratory paths for CGE-derived interneurons in the developing mouse brain. *Development* 143, 1753–1765.
- Wickham, H. (2009). *ggplot2: Elegant Graphics for Data Analysis* (Springer-Verlag New York).
- Yang, X., Cao, P., and Südhof, T.C. (2013). Deconstructing complexin function in activating and clamping Ca<sup>2+</sup>-triggered exocytosis by comparing knockout and knockdown phenotypes. *Proc. Natl. Acad. Sci. USA* 110, 20777–20782.
- Yu, F.H., and Catterall, W.A. (2004). The VGL-chanome: a protein superfamily specialized for electrical signaling and ionic homeostasis. *Sci. STKE* 2004, re15.
- Zeisel, A., Muñoz-Manchado, A.B., Codeluppi, S., Lönnberg, P., La Manno, G., Juréus, A., Marques, S., Munguba, H., He, L., Betsholtz, C., et al. (2015). Brain structure. Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. *Science* 347, 1138–1142.

## STAR★METHODS

## KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
<b>Antibodies</b>		
IF: chicken anti-RFP	Rockland	Cat# 600-901-379; RRID: AB_10704808
IF: rabbit anti-RFP	Rockland	Cat# 600-401-379; RRID: AB_2209751
IF: anti-CASPR (Clone K65/35)	EMD Millipore	Cat# MABN69; RRID: AB_10806491
IF: anti-Sst	Peninsula laboratories	Cat# T-4103.0050; RRID: AB_518614
IF: anti-Nos1	Thermo Fisher	Cat# 61-7000; RRID: AB_2533937
<b>Chemicals, Peptides, and Recombinant Proteins</b>		
ERCC RNA Spike-In Control Mixes	Thermo Fisher	Cat# 4456740
SuperScript III	Thermo Fisher	Cat# 18080093
RNaseOUT Recombinant Ribonuclease Inhibitor	Thermo Fisher	Cat# 10777019
RNA fragmentation buffer	New England Biolabs	Cat# E6105S
RNA MinElute kit	QIAGEN	Cat# 74204
Antarctic phosphatase	New England Biolabs	Cat# M0289
Poly nucleotide kinase	New England Biolabs	Cat# M0201
T4 RNA ligase2, truncated	New England Biolabs	Cat# M0242
Ampure XP beads	Beckman Coulter	Cat# A63880
SPRIselect size selection beads	Thermo Fisher	Cat# B23317
DL-AP5	Tocris	Cat# 0105
CNQX	Tocris	Cat# 1045
TTX	Tocris	Cat# 1078
Protease from Streptomyces griseus	Sigma-Aldrich	Cat# P5147
<b>Critical Commercial Assays</b>		
Message Amp II kit	Thermo Fisher	Cat# AM1751
Illumina TrueSeq smallRNA kit	Illumina	Cat# RS-200-0012
Bioanalyzer RNA Pico chip	Agilent	Cat# 5067-1513
Bioanalyzer High Sensitivity DNA chip	Agilent	Cat# 5067-4626
Panomics ViewRNA ISH Tissue 2-Plex Assay kit	Affymetrix	Cat# QVT0012
<b>Deposited Data</b>		
Raw scRNA-seq data	This paper	GEO: GSE92522
<b>Experimental Models: Organisms/Strains</b>		
Mouse: Nkx2-1 tm1.1(cre/ERT2)Zjh/J	Jackson laboratory	Stock No. 014552
Mouse: Nkx2-1 tm2.1(flpo)Zjh/J	Jackson laboratory	Stock No: 028577
Mouse: B6.129P2-Pvalb tm1(cre)Arbr/J	Jackson laboratory	Stock No: 017320
Mouse: Sst tm3.1(flpo)Zjh/J	Jackson laboratory	Stock No: 028579
Mouse: B6;129S-Nos1 tm1.1(cre/ERT2)Zjh/J	Jackson laboratory	Stock No: 014541
Mouse: B6(Cg)-Calb2 tm2.1(cre/ERT2)Zjh/J	Jackson laboratory	Stock No: 013730
Mouse: Vip tm2.1(flpo)Zjh/J	Jackson laboratory	Stock No: 028578
Mouse: Cck tm1.1(cre)Zjh/J	Jackson laboratory	Stock No: 012706
Mouse: Ptn-CreER	This paper	N/A
<b>Oligonucleotides</b>		
N10B1: CGATTGAGGCCGGTAATACGACTCACTATAGG GGTTCAGAGTTCTACAGTCGACGATCNNNNNNNNNN GTAACCACTTTTTTTTTTTTTTTTTTTTTTTTTVN	This paper	N/A
N10B2: CGATTGAGGCCGGTAATACGACTCACTATAGG GGTTCAGAGTTCTACAGTCGACGATCNNNNNNNNNN TAGGAGCACTTTTTTTTTTTTTTTTTTTTTTVN	This paper	N/A

(Continued on next page)

***Continued***

REAGENT or RESOURCE	SOURCE	IDENTIFIER
N10B3: CGATTGAGGCCGGTAATACGACTCACTATAGG GGTCAGAGTTCTACAGTCCGACGATCNNNNNNNNN GCACAGGACTTTTTTTTTTTTTTTTTTTTVN	This paper	N/A
N10B4: CGATTGAGGCCGGTAATACGACTCACTATAGG GGTCAGAGTTCTACAGTCCGACGATCNNNNNNNNN CTCTGGACTTTTTTTTTTTTTTTTTTVN	This paper	N/A
N10B5: CGATTGAGGCCGGTAATACGACTCACTATAGG GGTCAGAGTTCTACAGTCCGACGATCNNNNNNNNN CCTGACGACTTTTTTTTTTTTTTTTTTVN	This paper	N/A
N10B6: CGATTGAGGCCGGTAATACGACTCACTATAGG GGTCAGAGTTCTACAGTCCGACGATCNNNNNNNNN GACGTGGACTTTTTTTTTTTTTTTTVN	This paper	N/A
N10B7: CGATTGAGGCCGGTAATACGACTCACTATAGG GGTCAGAGTTCTACAGTCCGACGATCNNNNNNNNN CTTCTGCACTTTTTTTTTTTTTTTTVN	This paper	N/A
N10B8: CGATTGAGGCCGGTAATACGACTCACTATAGG GGTCAGAGTTCTACAGTCCGACGATCNNNNNNNNN NAAGATCGACTTTTTTTTTTTTTTTTVN	This paper	N/A
N10B9: CGATTGAGGCCGGTAATACGACTCACTATAGG GGTCAGAGTTCTACAGTCCGACGATCNNNNNNNNNC GTGACCACCTTTTTTTTTTTTTTTTVN	This paper	N/A
N10B10: CGATTGAGGCCGGTAATACGACTCACTATAG GGGTTCAGAGTTCTACAGTCCGACGATCNNNNNNNNN NCTTATGGACTTTTTTTTTTTTTTTTVN	This paper	N/A
N10B11: CGATTGAGGCCGGTAATACGACTCACTATAG GGGTTCAGAGTTCTACAGTCCGACGATCNNNNNNNNN NTGGTACCACTTTTTTTTTTTTTTTTVN	This paper	N/A
N10B12: CGATTGAGGCCGGTAATACGACTCACTATAG GGGTTCAGAGTTCTACAGTCCGACGATCNNNNNNNNN NGTCTACGACTTTTTTTTTTTTTTTTVN	This paper	N/A
N10B13: CGATTGAGGCCGGTAATACGACTCACTATAG GGGTTCAGAGTTCTACAGTCCGACGATCNNNNNNNNN NNCCTAAGCACTTTTTTTTTTTTTTTTVN	This paper	N/A
N10B14: CGATTGAGGCCGGTAATACGACTCACTATAG GGGTTCAGAGTTCTACAGTCCGACGATCNNNNNNNNN NCGCGTCGACTTTTTTTTTTTTTTTTVN	This paper	N/A
N10B15: CGATTGAGGCCGGTAATACGACTCACTATAG GGGTTCAGAGTTCTACAGTCCGACGATCNNNNNNNNN NNCAGTTGCACTTTTTTTTTTTTTTVN	This paper	N/A
N10B16: CGATTGAGGCCGGTAATACGACTCACTATAG GGGTTCAGAGTTCTACAGTCCGACGATCNNNNNNNNN NAGCAAGCACTTTTTTTTTTTTTTTTVN	This paper	N/A
T7-RA5: CGATTGAGGCCGGTAATACGACTCACTATAG GGGTTCAGAGTTCTACAGTCCGACGATC	This paper	N/A
Second round primer: NNNNNN	This paper	N/A
<b>Software and Algorithms</b>		
Bowtie2	Langmead and Salzberg, 2012	<a href="http://bowtie-bio.sourceforge.net/bowtie2/index.shtml">http://bowtie-bio.sourceforge.net/bowtie2/index.shtml</a>
Samtools	Li et al., 2009	<a href="http://samtools.sourceforge.net/">http://samtools.sourceforge.net/</a>
Custom Python script to count tags	Crow et al. 2016	<a href="https://github.com/maggiecrow/scCoexp">https://github.com/maggiecrow/scCoexp</a>
MetaNeighbor in R	Crow et al. 2017; <a href="http://biorxiv.org/lookup/doi/10.1101/150524">http://biorxiv.org/lookup/doi/10.1101/150524</a>	<a href="https://github.com/maggiecrow/MetaNeighbor">https://github.com/maggiecrow/MetaNeighbor</a>

(Continued on next page)

**Continued**

REAGENT or RESOURCE	SOURCE	IDENTIFIER
R and RStudio	The R Foundation	<a href="https://www.r-project.org/">https://www.r-project.org/</a> ; <a href="https://www.rstudio.com/">https://www.rstudio.com/</a>
Fiji	Schindelin et al. 2012	<a href="https://fiji.sc">https://fiji.sc</a>
Velocity Software	Perkin Elmer	<a href="http://cellularimaging.perkinelmer.com/downloads/">http://cellularimaging.perkinelmer.com/downloads/</a>
SoftWoRx 6.5.2 software	GE Healthcare	<a href="http://incelldownload.gehealthcare.com/bin/download_data/SoftWoRx/6.5.2/SoftWoRx.htm">http://incelldownload.gehealthcare.com/bin/download_data/SoftWoRx/6.5.2/SoftWoRx.htm</a>
Imaris 7.6.5.	Bitplane	<a href="http://www.bitplane.com/releasenotes.aspx">http://www.bitplane.com/releasenotes.aspx</a>
Other		
FISH ViewRNA probe: Cdh1 NM_009864.2, type1	Affymetrix	Cat# VB1-15084
FISH ViewRNA probe: Gpr88 NM_022427.2, type1	Affymetrix	Cat# VB1-13597
FISH ViewRNA probe: Hapl1 NM_013500, type6	Affymetrix	Cat# VB6-13436
FISH ViewRNA probe: Hhip NM_020259.4, type1	Affymetrix	Cat# VB1-18918
FISH ViewRNA probe: Htr7 NM_008315.2, type1	Affymetrix	Cat# VB1-15758
FISH ViewRNA probe: Pthlh NM_008970.3.4, type1	Affymetrix	Cat# VB1-12560
FISH ViewRNA probe: Ptn NM_008973.2, type1	Affymetrix	Cat# VB1-18915
FISH ViewRNA probe: Slc7a3 NM_007515.3, type1	Affymetrix	Cat# VB1-18914
FISH ViewRNA probe: Tacr1 NM_009313.5, type1	Affymetrix	Cat# VB1-12918
FISH ViewRNA probe: Unc5b NM_029770, type6	Affymetrix	Cat# VB6-13432
FISH ViewRNA probe: Td-Tomato, type6	Affymetrix	Cat# VB6-13925
FISH ViewRNA probe: Td-Tomato, type1	Affymetrix	Cat# VF1-14985

**CONTACT FOR REAGENT AND RESOURCE SHARING**

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Dr. Z. Josh Huang ([huangj@cshl.edu](mailto:huangj@cshl.edu)).

**EXPERIMENTAL MODEL AND SUBJECT DETAILS****Animals**

Nkx2.1-CreER, Pv-ires-Cre animals were bred separately to Ai14 reporter to label CHC and PVBC in the cortex respectively. CHCs were enriched in frontal cortex with tamoxifen induction at E17.5. Intersectional labeling of PCPs were achieved by breeding (a) Sst-Flp, Nos1-CreER for LPN, (b) Sst-Flp, CR-Cre, for MNC (c) VIP-Flp, CR-Cre for ISC and (d) VIP-Flp, CCK-Cre for CCKC separately to Ai65 intersectional reporter that will label cells with tdTomato only when both the lox-Stop-lox and Frt-STOP-Frt cassettes are excised (Figure S1). For validation studies mice were also bred to Nkx-ires-FlpO (He et al., 2016) to get Nkx2.1-Flp; Cck-Cre; Ai65. Ptn-CreER animals were generated as described in Figure S6B and were induced postnatal at P5. Mice were bred and maintained according to animal husbandry protocols at Cold Spring Harbor Laboratory (Institutional Animal Care and Use Committee reference number 16-13-09-8) with access to food and water ad libitum and 12 hr light-dark cycle.

**METHOD DETAILS****Phenotype characterized GABAergic subpopulations**

Cortical GABAergic neurons can be parsed into several non-overlapping populations and, in a few cases, bona-fide types based on developmental origin, innervation targets, and molecular markers (Kepecs and Fishell, 2014). The embryonic medial and caudal ganglionic eminences (MGE and CGE) give rise to two broad groups, the former is divided into parvalbumin (PV) and somatostatin (SST) populations and the latter is marked by HTR3a (Kepecs and Fishell, 2014) (Figures 1A and 1B). The PV population includes fast-spiking basket cells (PVBC) that innervate the perisomatic region (Hu et al., 2014) and chandelier cells (CHC) that target the axon initial segment (AIS) (Taniguchi et al., 2013). The SST population includes Martinotti cells (MNC) that target distal dendrites (Silberberg and Markram, 2007), long projection cells (LPC) (Kilduff et al., 2011) and multiple other types. The HTR3a group includes the Vasoactive

intestinal peptide (VIP) and Reelin populations, and the VIP population comprises interneuron-selective dis-inhibitory cells (ISC) (Staiger et al., 2004), Cholecystokinin (CCK) small basket cells (CCKC) (Armstrong and Soltesz, 2012) and likely additional types. Accumulated anatomical, physiological, and molecular evidence indicate that these are non-overlapping subpopulations, and CHC, LPC and PVBC are considered cardinal types (He et al., 2016).

### **RNA double in situ and imaging**

RNA double in situ was performed using Quantigene ViewRNA tissue ISH (Affymetrix, USA) following manufacturer's recommended protocol. Fresh unfixed brain tissues were frozen in OCT blocks using dry-ice isopentane slurry. Brains can be stored in -80°C until cryosectioning. Cryosectioning was done on Leica cryotome at 12μm thickness, and sections collected on charged glass slides. Custom and off-the shelf branched-DNA oligo ISH probes were designed and synthesized by Affymetrix Quantigene ViewRNA. Sections on slides were postfixed just prior to ISH, and in situ steps were followed according to manufacturer's recommended protocol. For dual signal detection QuantiGene Type-1 and Type-6 probes were used. Fluorescent signals from Type-1 and Type-6 ISH probes were imaged on tile-scanning mode using Perkin Elmer spinning disk confocal at 10X magnification and auto-stitched using Velocity software. Stitched images were exported as TIFFs for further processing and adjustments to brightness and contrast in FIJI (Fiji is just ImageJ) and assembled in Adobe Illustrator.

### **Super-resolution microscopy**

Longitudinal brain sections from Sst-Flp;Nos1-Cre;Ai65 animals were perfused and sectioned at 75μm thickness then immunolabeled with anti-Caspr (EMD Millipore #MABN69, 1:500 dilution) and anti-RFP (Rockland #600-401-379, 1:1000 dilution). Super resolution images were acquired with GE Healthcare OMX V3 structured illumination microscopy system using: 488 and 593 nm solid state lasers; UPlanS Apochromat 100 × 1.4 NA objective lens (Olympus); 2 EM-CCD cameras (Cascade II 512, Photometrics). 3D structured illumination images were reconstructed with SoftWoRx® 6.5.2 software. 3D rendering was performed using Imaris (Bitplane) 7.6.5. exported as TIFF, processed in FIJI and assembled in Adobe Illustrator.

### **Manual cell sorting**

To isolate individual RFP-labeled GABAergic neurons, we microdissected motor and somatosensory cortical slices from fresh brain tissues of mature (6 weeks old) mice, generated single cell suspension and manually purified single RFP-labeled cells (Sugino et al., 2006). Brains were sectioned at 300 μm thickness using a cooled stage vibratome (Microm, Model HM360) with circulating oxygenated artificial cerebrospinal fluid. Sections were blocked in AP5, CNQX, and TTX cocktail to prevent excitotoxic cell death and then treated with mild protease (Fraction IV protease Streptomyces, Sigma Cat#P5147-5G). Brain regions of interest were microdissected and triturated to dissociate the cells. Dissociated cells were put into a Petri dish in low density for optimal cell-cell separation then purified progressively by transferring RFP cells to fresh plates 3 times. Finally single RFP-positive cells was collected using patch pipette capillary and dispensed individually into separate single tubes pre-filled with RNaseOUT (Invitrogen), ERCC spike-in RNAs in 1:400 K dilution, sample specific RT primers for a total of 1 μL volume. Process was repeated to collect 32-64 cells in one manual cell sorting session. Cells were flash frozen in liquid nitrogen and stored at -80°C until processed. Patch pipette was single use only and fresh pipettes were used for every single cell collected. Manual sorting resulted in negligible contaminants as shown by glial and excitatory neuronal transcripts in Figure S1.

### **Linear RNA amplification**

Single cell mRNAs were converted to cDNAs through polyA primers containing a sample barcode and unique molecular identifiers (UMIs). We employed two rounds of in vitro transcription amplification (Eberwine et al., 1992) followed by Illumina TrueSeq protocol to construct RNAseq libraries (Hashimshony et al., 2012).

Custom T7-polyA primers (N10B1 → N10B16, Integrated DNA Technologies, USA) were designed containing 9bp error correcting sequences for identifying single cells (sample barcode) and 10bp random nucleotide sequences (UMI/varietal tag) to label each mRNA molecule amplified with a unique barcode. The UMI allows for elimination of reads containing duplicate tags for the same mapped sequence and only tally up the total unique tags of all mapped sequence to a coding sequence. This primer also contained a 26bp flanking RA5 adaptor sequence needed for downstream Illumina cDNA library step, which eliminates a rate limiting enzymatic 5' ligation step of cDNA preparation increasing efficiency (Hashimshony et al., 2012).

RNA was linearly amplified by T7 RNA polymerase using two rounds of in-vitro transcription (MessageAmp-II kit Life Technologies) according to the manufacturer's recommended protocol with some modifications. Cells were lysed by repeated heating to 70°C and snap cooling to 4°C and first strand synthesis was carried out at 42°C for 2hrs with first strand buffer, dNTP mix, RNase inhibitor and ArrayScript enzyme. Second strand synthesis was done at 16°C for 2hrs with second strand buffer, dNTP mix, T4 DNA polymerase and RNaseH. cDNA was purified using columns and first round IVT was performed at 37°C for 14hrs to make aRNA. For the 2nd round of linear amplification column purified aRNA from first IVT underwent another first strand synthesis at 42°C for 2hrs using second round primers, followed by RNaseH digestion at 37°C for 30mins and another 2nd round second strand synthesis at 16°C with a T7-RA5 primer. The resulting double stranded cDNA underwent a final second IVT step at 37°C for 14hrs to make aRNA. These two rounds of linearly amplified aRNA products now carried the 3' end of the polyA transcripts for mapping to coding regions plus the sample barcode to indicate which PCP it came from, UMI sequence for counting unique cell-endogenous parent mRNA

molecules and one of the flanking sequence (RA5 adaptor) for Illumina sequencing. Second round aRNAs were fragmented chemically using NEBNext® Magnesium RNA Fragmentation Module (Cat#E6150S), column purified using RNA MinElute (QIAGEN) for final Illumina cDNA library preparation steps.

### cDNA library generation and sequencing

cDNA library was generated using Illumina TruSeq small RNA kit (Cat#RS-200-0012) and only 3'-adaptor (RA3) need to be ligated enzymatically using truncated T4 RNA ligase (NEB M0242) on to the fragmented aRNA and the 5' ligation step for RA5-adaptor was skipped (Hashimshony et al., 2012). Adaptor ligated fragmented aRNA was reverse transcribed using SuperScriptIII reverse transcriptase (Invitrogen, USA) and PCR enriched using TruSeq indices (for multiplexing) for no more than 7–11 cycles. The resulting library was size-selected using SPRISelect magnetic beads (Agencourt) to select 350–450bp fragments and paired-end sequenced for 101bp in Illumina HiSeq. No more than 32 single cells were run in one lane of HiSeq2000 generating on average ~180–200 million reads per lane.

## QUANTIFICATION AND STATISTICAL ANALYSIS

### Mapping and tag counting

As in our previous work (Crow et al., 2017), Bowtie (v 0.12.7) was used for sequence alignment of read2 (polyA primed) to the mouse reference genome (mm9), while read1 sequences were used for UMI (varietal-tag) counting. A custom python script was used Bowtie (v 0.12.7) was used for sequence alignment of read2 (polyA primed) to the mouse reference genome (mm9), while read1 sequences were used for UMI (varietal-tag) counting. Multiple reads to the same gene with the same tag sequences were rejected and only counted as one, such that only mapped sequences with unique tags were retained and tallied for each mRNA for each cell.

We obtained  $\sim 4.8 \times 10^5$  (median, Avg =  $6.9 \times 10^5$ ) mapped reads per cell, each containing  $\sim 1.0 \times 10^5$  (median, Avg =  $1.4 \times 10^5$ ) unique reads that typically detected on average  $\sim 10,000$  genes (range  $\sim 7,500$  to  $\sim 12,000$  genes median), with > 95% of the single cells detecting > 6,000 genes (Figures S1B and S1C). In each single cell ERCC spike-in RNA (Life Technologies) were used as internal controls, for which the absolute number of molecules that are added to sample can be calculated; this gave a linear relationship of I/O measures with a slope of 0.92 and adjusted  $R^2 = 0.96$  (Figure S1D). Following quality control screen, we obtained high depth transcriptome of  $\sim 584$  cells from the 6 PCPs (Figure S1D). This unique dataset thus contains high-resolution transcriptomes of phenotype-defined cortical GABAergic PCPs.

For any given gene the absolute unique counts were normalized to the total unique counts across all genes in a single cell and are expressed as unique Transcripts Per Million (uTPM). To determine differential gene expression (DE) and calculate fold-change, gene-wise Fisher's meta-analytic p value was calculated on these normalized gene expression values without further batch effect correction.

Compared with previous UMI-based (Zeisel et al., 2015) (detects 1.8–4.7K genes/cell) and non-UMI based method (Tasic et al., 2016) (detects  $\sim 7.2$ K genes/cell), our linear amplification with 10bp UMIs improved gene detection and quantification ( $\sim 9$ K genes/cell) (Figure S1H). Compared with Dropseq which allows vast throughput at lower resolution (Shekhar et al., 2016), our complementary approach achieved more comprehensive and quantitative transcriptome measurement of targeted cell populations.

### Fisher's meta-analytic Differential Expression

To assess differential gene expression across PCPs, we took advantage of the replicate batches within each type and performed a meta-analysis across replicates based on non-parametric statistics. Briefly, for each PCP we performed one-tailed Mann-Whitney tests between individual batches within a cell type against all cells outside of that cell type. To ensure that significance would arise from replication rather than extreme p values, prior to meta-analysis with Fisher's method, p values at FDR  $\leq 0.05$  for an individual test were set to the maximum p value meeting that criterion. Finally, meta-analytic p values were FDR corrected. Differentially expressed gene sets were defined by FDR adjusted p value  $< 0.05$  having log<sub>2</sub> fold change  $> 2$ .

### MetaNeighbor

To measure PCP identity we use the MetaNeighbor method as described in our companion paper (Crow et al., 2017). In brief, MetaNeighbor requires the input of a set of genes, an expression matrix and two sets of labels: one set for labeling each experiment, and one set for labeling the cell types of interest. We perform a stratified cross-validation which allows us to explicitly block technical sources of variation in single-cell analysis, in close parallel to our meta-analytic evaluation of single-cell data (Crow et al., 2017). Here, each batch was treated as an “experiment,” and we aimed to measure the replicability of cell identity across batches. Cell-type labels are held back from one experiment at a time and then predicted based on the others, to determine which gene sets functionally characterize cells across technical variation. For each gene set being used to evaluate a given cell-type, the method generates a network based on the Spearman correlation between all cells across the genes within the set. The correlation is rank standardized to provide network weightings between each pair of cells, and then a neighbor voting predictor scores cells as possessing a given annotation. The score is calculated as the sum of a given cell's connectivity weighting to neighbors possessing a given cell annotation. The score is calculated as the sum of a given cell's connectivity weighting to neighbors possessing a given cell annotation. For cross-validation, we permute through all possible combinations of leave-one-batch-out cross-validation, and report

the degree to which cells of the same type are recovered as the mean area under the receiver operator characteristic curve (AUROC) across all folds. To improve speed, AUROCs are calculated analytically:

$$AUC_j = \left( \frac{\sum_{(i|Cell_i \in |Cell\_type_j)} Ranks_i - \frac{N_{Pos} * (N_{Pos} + 1)}{2}}{N_{Pos} * N_{Neg}} \right)$$

where “Ranks” are the ranks of the hidden positives,  $N_{pos}$  is the number of true positives, and  $N_{neg}$  is the number of true negatives.

### Notched Boxplots

Notched boxplots shown in main [figures 3–7](#) and supplement [figures S1, S3–S5](#): box represent interquartile range (IQR) 50 percent of data, notch is 95% confidence interval of the median, horizontal line inside box is median, upper whisker is lesser of 75 percentile or maximum value, lower whisker is greater of 25 percentile or minimum value, dots show outliers. According to Graphical Methods for Data Analysis ([Chambers, 1983](#)) although not a formal test the, if two boxes’ notches do not overlap there is ‘strong evidence’ (95% confidence) their medians differ. Package “ggplot2” (<https://cran.r-project.org/web/packages/ggplot2/ggplot2.pdf>) was used to plot notched boxplots ([Wickham, 2009](#))

### DATA AND SOFTWARE AVAILABILITY

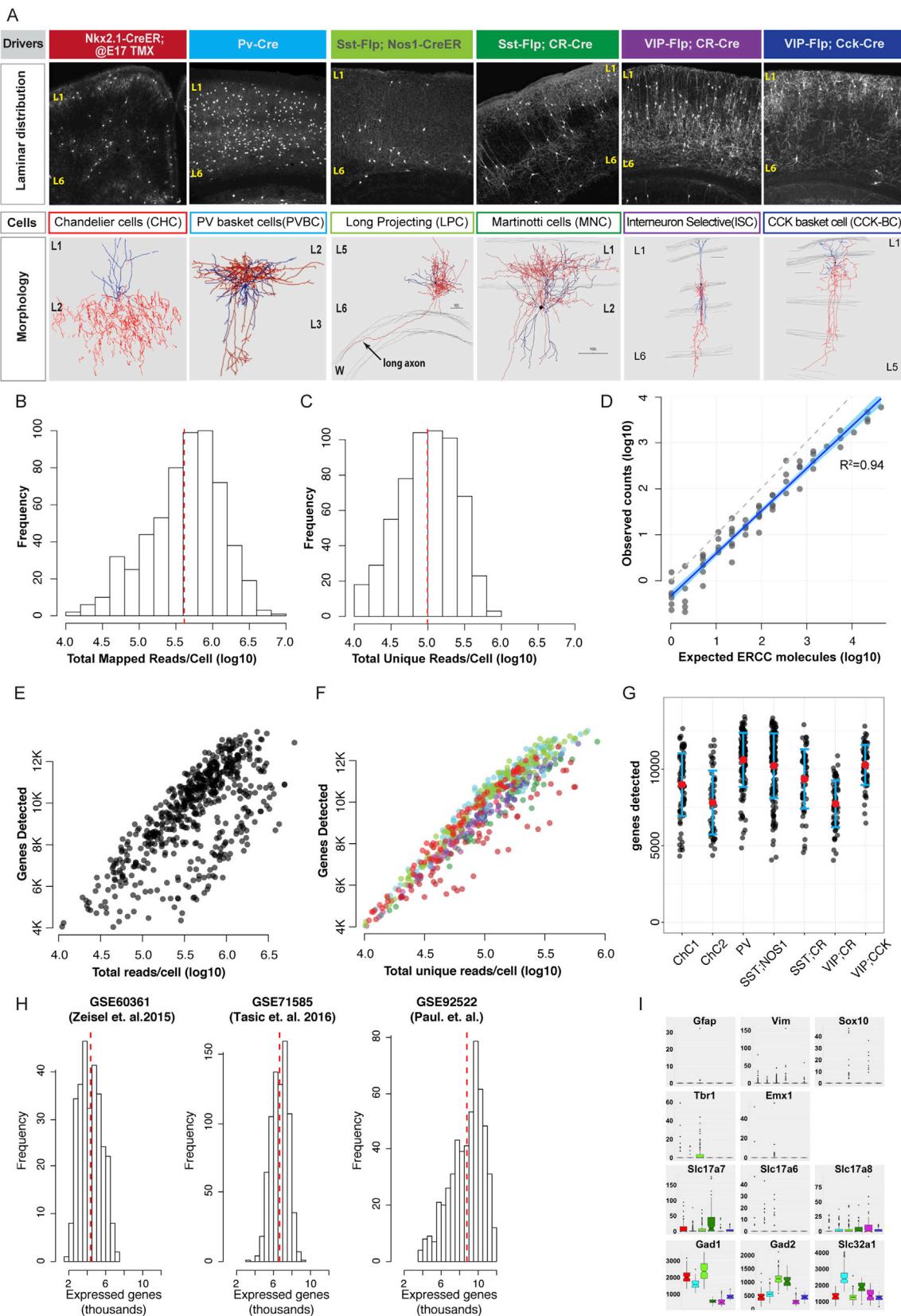
A Github repository containing R scripts and parsed data can be found online (see [Key Resources Table](#)). Raw data files, parsed data, and metadata have been uploaded to GEO: GSE92522.

### Ethics

Mice were bred and maintained according to animal husbandry protocols at Cold Spring Harbor Laboratory (Institutional Animal Care and Use Committee reference number 16-13-09-8).

# Supplemental Figures

Cell

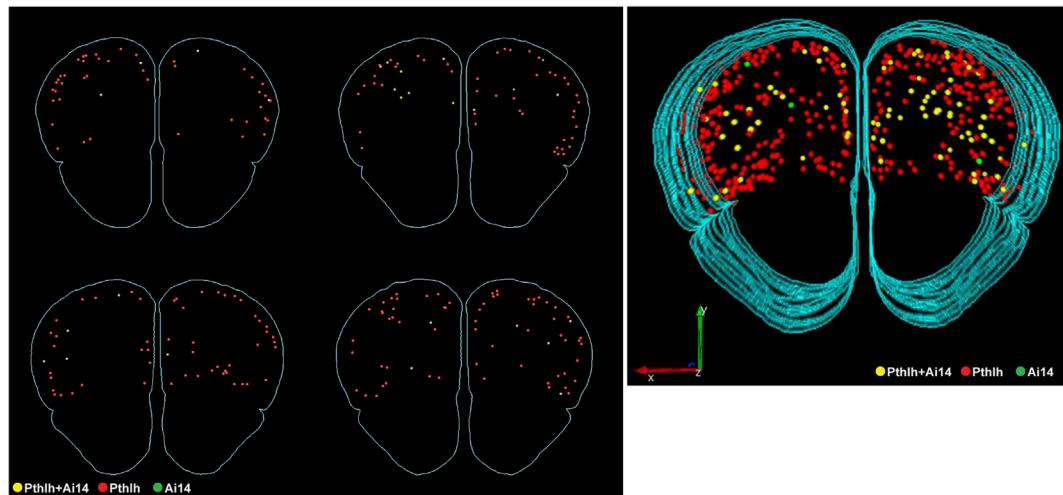
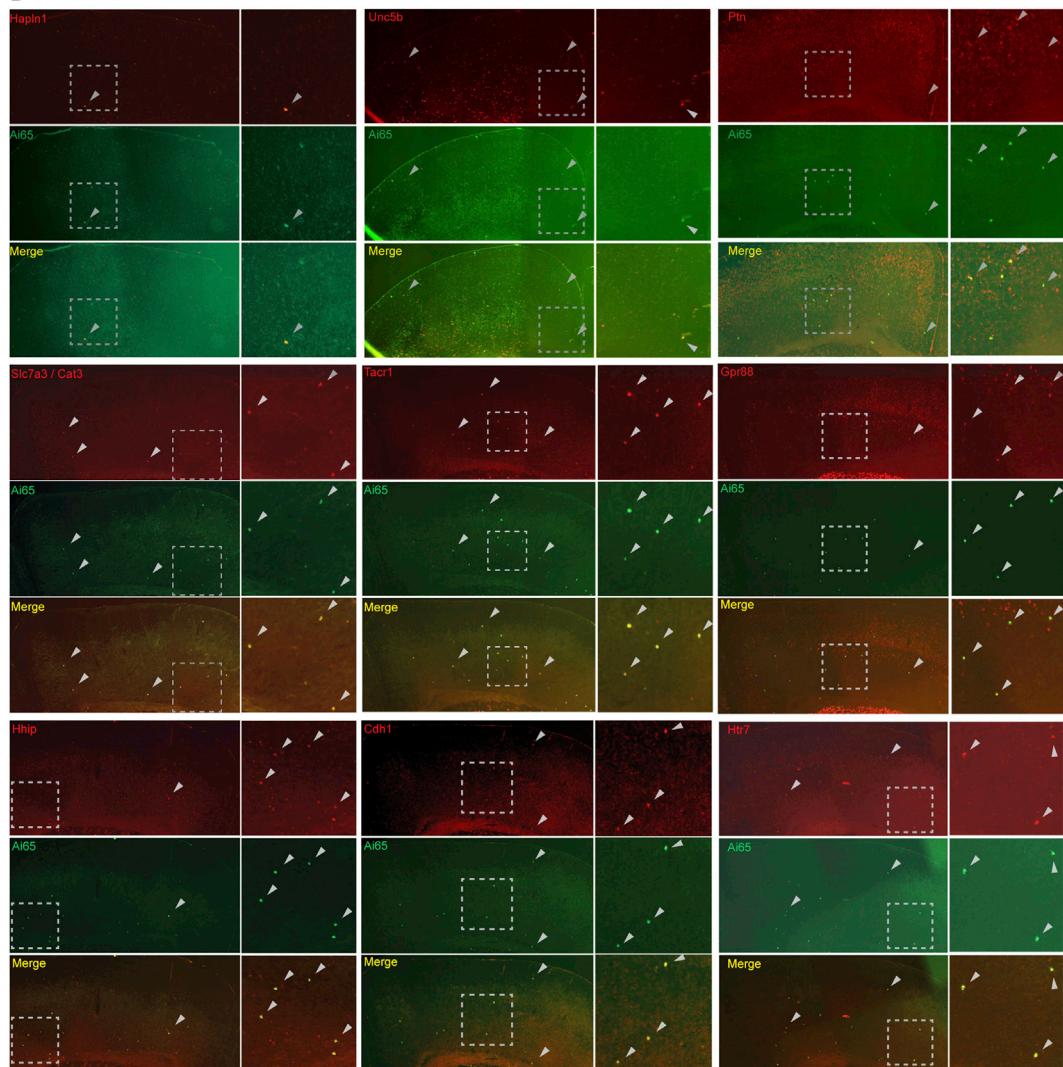


(legend on next page)

---

**Figure S1. Transcriptomic Analysis of Cortical GABAergic Phenotype-Characterized Populations, Related to Figure 1**

- (A) Intersectional, lineage and birth time dependent labeling (top row) of 6 PCPs with characteristic laminar distribution (middle row). Representative single cell reconstruction depicted characteristic morphology in each PCP (bottom row).
- (B) Histogram of total mapped reads per single cell with a median read depth of  $5.7 \times 10^5$  reads per cell.
- (C) Histogram of unique reads per single cell with a median of  $5.0 \times 10^5$  counts.
- (D) Plot of ERCC observed unique reads versus the numbers of molecules expected for each species from the ERCC cocktail shows linearity and slope close to 1; unity line shown as dotted gray line, blue shaded region shows 95% confidence interval.
- (E) Number of genes detected in single cells versus total mapped reads
- (F) Unique reads shows that gene detectability is correlated to read counts but there are no gene detection bias toward any of the PCPs (color code as in A).
- (G) Genes detected across PCPs range between 7.5 – 12 thousands (median).
- (H) Higher levels of genes detected in single cells compared to GSE60361 ([Zeisel et al., 2015](#)) and GE71585 ([Tasic et al., 2016](#)) dataset.
- (I) Negligible glial transcripts observed; such as Gfap, Vim, Sox10 and excitatory neuronal mRNAs such as Tbr1, Emx1, Slc17a7, Slc17a6 and Slc17a8. In contrast, high expression of typical GABAergic observed transcripts such as Gad1, Gad2 and Slc32a1.
- See also [STAR Methods, Notched boxplots](#).

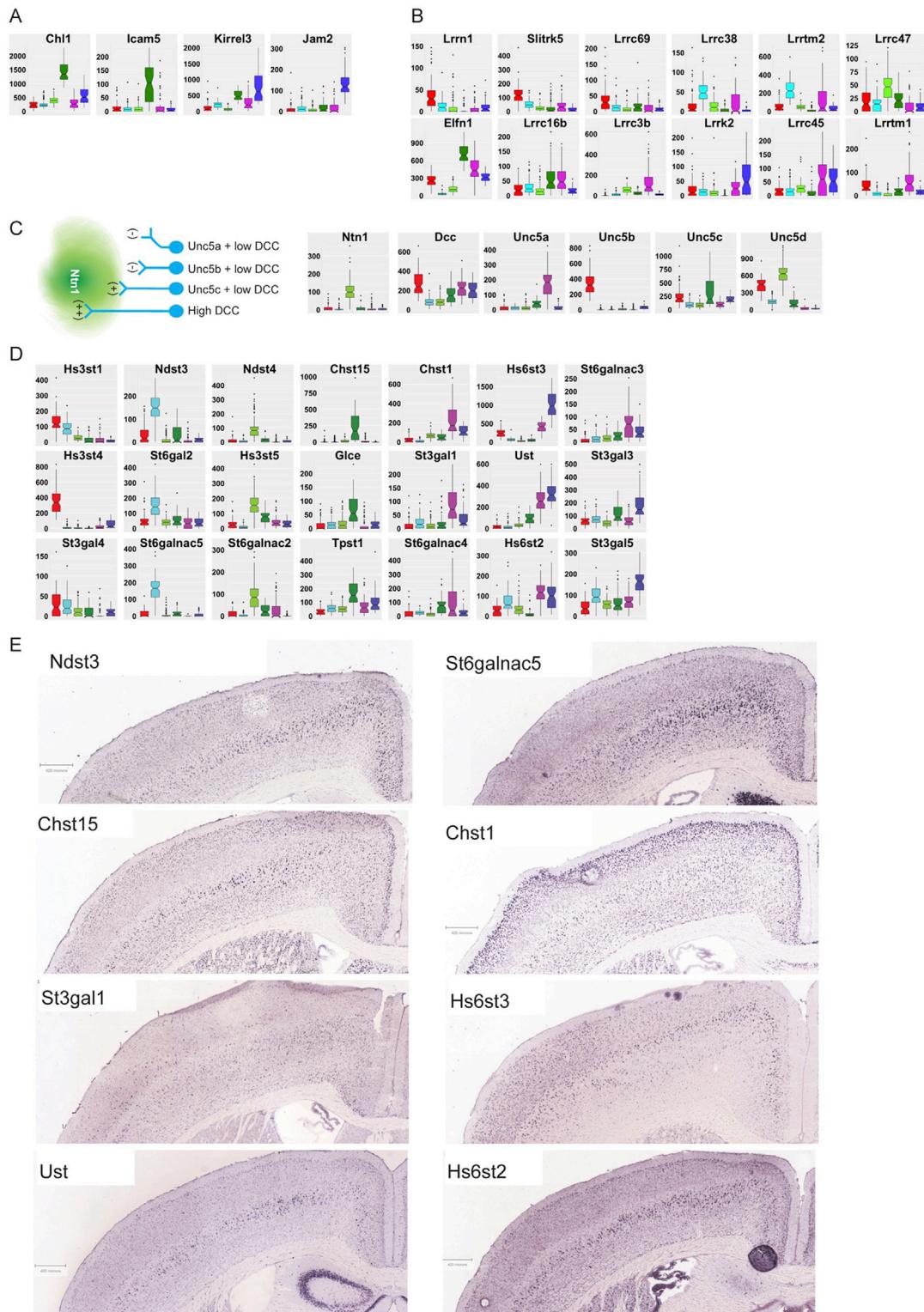
**A****B**

(legend on next page)

**Figure S2. Fluorescent Double In Situ in Brain Tissue Confirms Co-localization of Select Enriched Transcripts in PCPs, Related to Figure 1**

(A) Top left: Representative sections from 3D render. 136/143 tdTomato<sup>+</sup> (~95%) co-express Pthlh. Laminar distribution of Pthlh signal was similar to those reported for CHC distribution pattern (Taniguchi et al., 2013). Right: Partial 3D reconstruction render from serial coronal sections (spanning 288um, rostro-caudal) of mouse forebrain showing double mRNA in situ of CHC enriched transcript Pthlh (red), Ai14/tdTomato (green) and co-localization (yellow). Only a subset of CHCs were labeled in *Nkx2.1-CreER;Ai14* mice by tamoxifen induction at E17.

(B) Representative single images from double RNA in situ of Ai14/tdTomato (green) shows co-localization (yellow) with select transcripts enriched in CHC2 (Unc5b, Hapln1) SST;NOS1 cells (Ptn, Slc7a3, Tacr1, Gpr88, Hhip, Cdh1 and Hrt7). Dotted box represents area in higher magnification, arrowheads indicate co-localization.

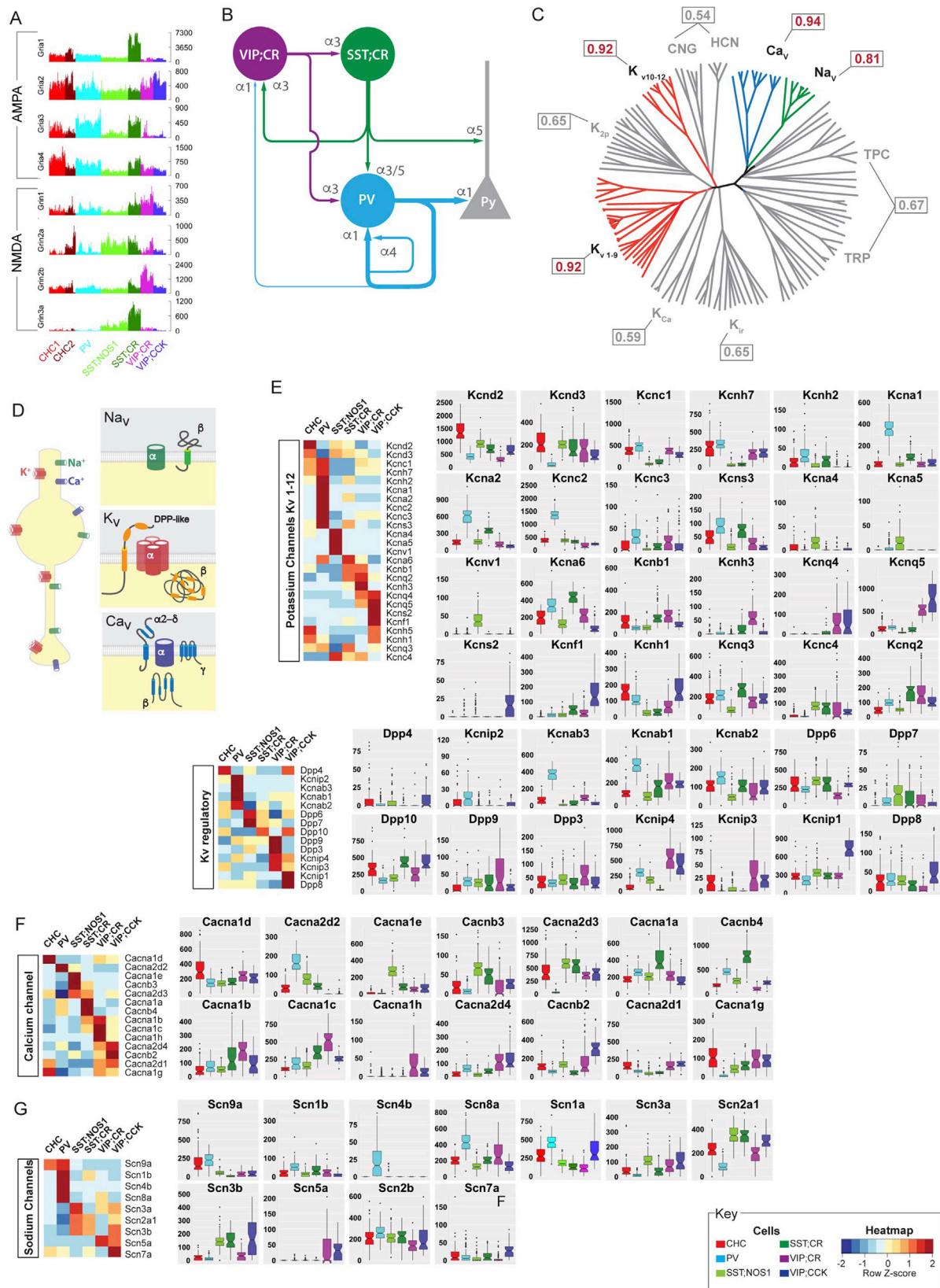


**Figure S3. Differential Expression of Netrin-unc5 and Carbohydrate-Modifying Enzymes among PCPs, Related to Figure 3**

(A) Approximately 20 immunoglobulin cell adhesion molecule are differentially expressed (also see Figure 3F) and may contribute to the cellular, subcellular and synaptic specificity of PCPs. Among them, CHL1 is particularly enriched in SST/CR population which includes dendrite-targeting Martinotti cells, consistent with its role in regulating subcellular synapse specificity; Kirrel, JAM2, 3, ICAM5 are each enriched in a specific subpopulation (also see Figure 3F).

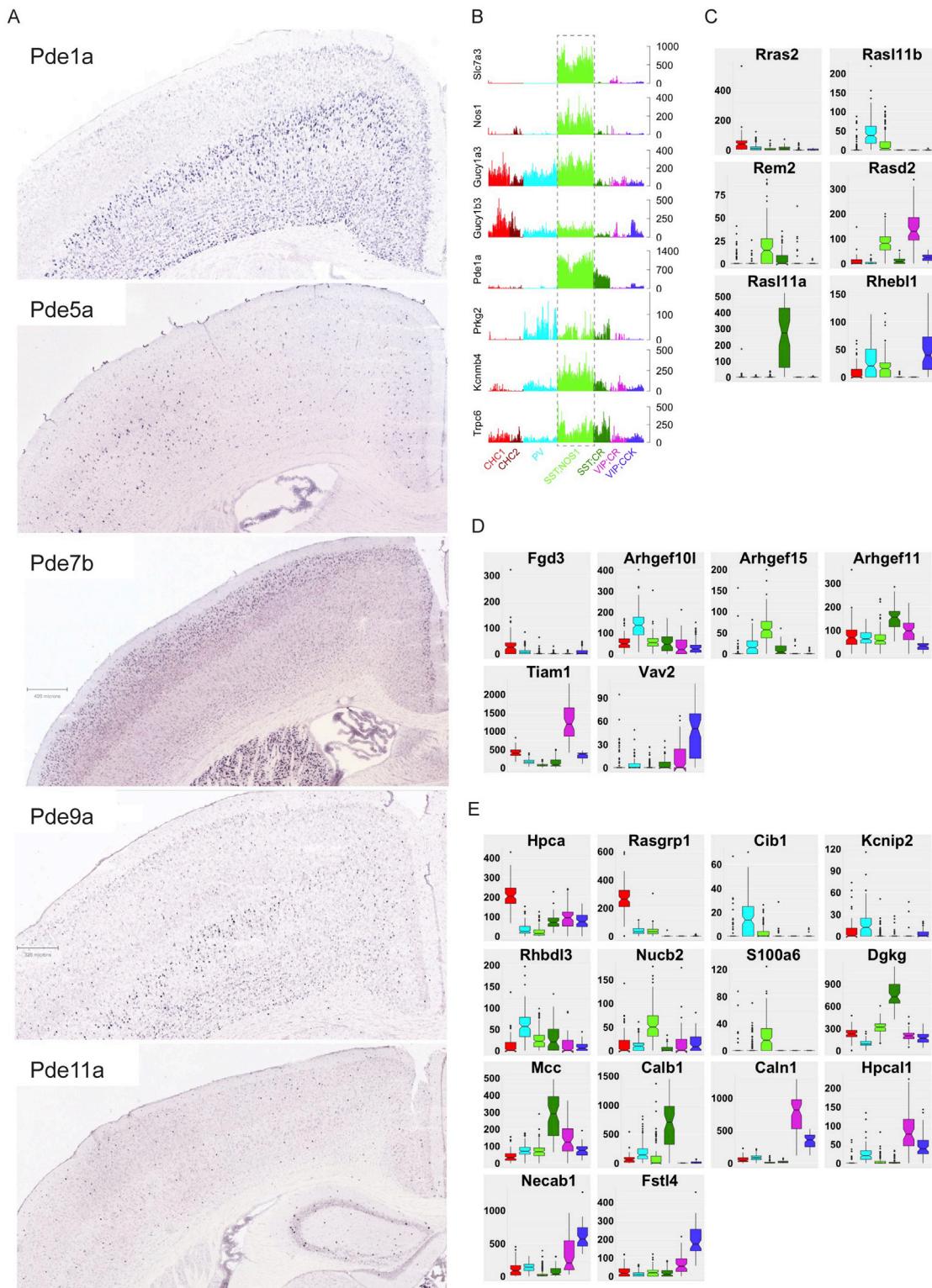
(legend continued on next page)

- 
- (B) Each PCP enriches for a different set of 6-12 LRR proteins (also see [Figure 3F](#)). Among these, Elfn1 is prominent in SST/CR cells (mostly Martinotti cells), cortical homolog of hippocampal O-LM interneurons in which Elfn1 contributes to the synaptic facilitation of their glutamatergic inputs
- (C) Schematic depiction of known attractive (+) and repulsive (-) interactions between the ligand Netrin (Ntn1) and its receptors - Unc5 family members and DCC. Boxplots show highly specific expression of Ntn1 and Unc5 family members among PCPs.
- (D) Boxplots of CAM modifying enzymes Sulfo- and Sialyl-transferases enriched in individual PCPs. Six sulfotransferase are highly specific to different PCPs: Hs3st4 to ChC, Chst15 to SST/CR, Chst1 to VIP/CR, Hs3st5 to SST/NOS; Hs3st1 is enriched in ChC and PV cells and Hs6st3 is enriched in VIP/CCK cells.
- (E) Representative *in situ* images from Allen Brain Atlas showing distinct laminar distribution patterns of some carbohydrate modifying enzymes in coronal brain sections. Many of these patterns are characteristic to the distribution of subsets of GABAergic neurons.
- See also [STAR Methods](#), [Notched boxplots](#).



(legend on next page)

- Figure S4. Differential Expression of Transmitter Receptors and Voltage-Gated Ion Channels among PCPs, Related to Figures 2 and 4**
- (A) CGE-derived VIP cells have overall relatively low AMPARs and roughly similar GluA1 and GluA2 levels (GluA1:GluA2 = 1.4), and VIP/CR cells have relatively more NMDARs especially those containing GluN2B (GluN2B:GluN2A = 11.0). MGE-derived cells have much higher levels of GluA1 (average GluA1:GluA2 = 8.4), with striking cell type differences: GluA1:GluA2 ranges from 4.1 in SST/NOS1 cells to 20.4 in SST/CR cells. While PV cells have highest GluA3 levels and CHCs have highest GluA4 levels, SST/CR cells show highest levels of GluA1 and highest non-GluA2/GluA2 ratio (24.8). Interestingly, SST/CR cells also have relatively high GluN2A:GluN2B ratio and high level GluN3A for NMDARs among the PCPs.
- (B) Distinct GABA<sub>A</sub>R subtypes with specific subunit combinations are likely targeted to specific connections that match the presynaptic terminals to optimize inhibitory transmission properties. PVBCs predominantly mediate self-inhibition (i.e., other PV cells) in addition to perisomatic inhibition of Py. They further receive inhibitory inputs from SST cells (mainly SST/CR positive Martinotti cells) and interneuron-selective VIP cells (mostly VIP/CR positive). PV-PV transmission is one of the fastest in the brain, mediated by the  $\alpha 1\beta 2\gamma 2$  subtype. It thus can be inferred that  $\gamma 3$ -containing slow kinetics receptors, likely abundant in PVBCs, are excluded from PV-PV synapses. The unique co-expression  $\alpha 4$  and  $\delta$ , a well-established combination for extrasynaptic and axonal GABA<sub>A</sub>R, suggest the presence of this subtype in PVBC terminals, likely activated by GABA spill-over during concerted GABA release from dense perisomatic synapses characteristic to PV axon terminals. These considerations further raise the possibility that other subunit combinations, such as those containing  $\alpha 3$ ,  $\alpha 5$ , and  $\gamma 3$ , might support inputs from SST/CR and VIP/CR cells, especially in the dendritic compartment of PVBCs. Following similar logic, SST/CR cells might receive VIP cell inputs through  $\alpha 3\beta 1/3\gamma 3$  type GABA<sub>A</sub>Rs; VIP cells likely receive PV cell input through  $\alpha 1$ -containing GABA<sub>A</sub>Rs and Martinotti cell input through  $\alpha 3$ -containing GABA<sub>A</sub>Rs.
- (C) An unrooted cladistic representation drawn using minimum evolution analysis of 143 members of structurally related voltage-gated ion channels based on their amino acid sequence of the minimal pore regions adopted without structural manipulations (modified from Yu and Catterall, 2004). Only three of these families, Kv, NaV and CaV are highly discriminative of PCPs as shown by their AUROC scores (boxed, red text shows high scoring families).
- (D) Schematic of the subcellular distribution of voltage gated ion channels and their subunit compositions of pore forming and accessory regulatory proteins.
- (E) Heatmap and selected boxplots showing differential expression of Kv channels and their regulatory subunits among PCPs. Fast-spiking PV basket cells express the highest levels and diversity of Kv channels.
- (F) Heatmap and selected boxplots showing differential expression of CaV channels.
- (G) Heatmap and selected boxplots showing differential expression of NaV channels. Fast-spiking PV basket cells express the highest levels and diversity of Nav channels. The relevance of ion channel transcription profiles to physiological properties is highlighted by the striking case of PVBCs. PVBCs convert an excitatory input to an inhibitory output within a millisecond, a stunning cell biology feat that appears to involve optimizing multiple aspects of electrical signaling across subcellular compartments, in part through specific expression and localization of a unique assortment of VGICs with highly tailored biophysical properties. The prominent enrichment of multiple Navs (Scn9a, Scn8a, Scn1a, Scn3a, Scn1b, Scn4b, Scn2b) likely underlie the “supercritical density” of Nav for ensuring fast signaling in PV cell axons, and the striking elevation of a large set of fast-kinetics Kv1-4 members may implement rapid repolarization and narrow AP duration at each subcellular domain. Expression of auxiliary subunits Kca, Kir and K2p further hints uncharacterized physiological properties.
- See also STAR Methods, Notched boxplots.



**Figure S5. Differential Expression of Regulatory Components of Intracellular Signaling Molecules among PCPs, Related to Figure 5**

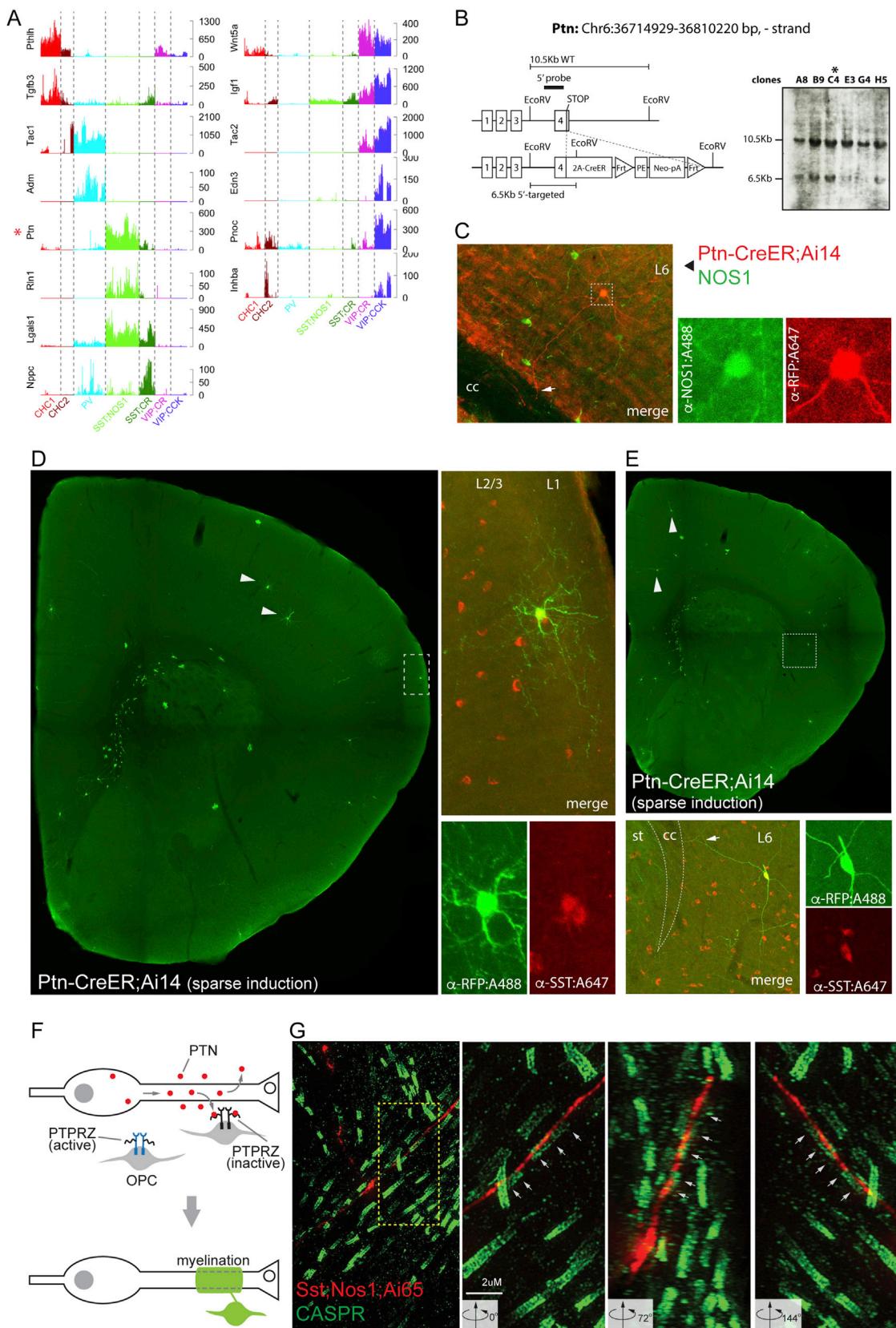
(A) Selected RNA in situ images from Allen Brain showing that multiple phosphodiesterases are expressed in restricted neuronal populations in adult mouse cortex with patterns characteristic to GABAergic interneurons.

(legend continued on next page)

(B) Barplots showing coordinated expression of multiple genes in SST/NOS1 cells encoding almost the entire NO-cGMP pathway. These include the substrate L-arginine transporter (*Slc7a3*), synthetic enzyme (*NOS1*), second messenger synthesis (*Gucy1 $\alpha$ 3* and *Gucy1 $\beta$ 3*), degradation (*Pde1a*) and signaling (*Prkg*). Furthermore, two *Prkg* targets, members of the transient receptor potential channels (*Trpc5*) and BK-type potassium channels ( $\alpha$ 1 core subunit and  $\beta$ 4 auxiliary subunits of *Kcnma*) are also differentially enriched.

(C–E) Boxplots for Ras (B), Rho-GEF (C), and EF-hand/CaBP (D) family members among PCPs.

See also [STAR Methods](#), [Notched boxplots](#).

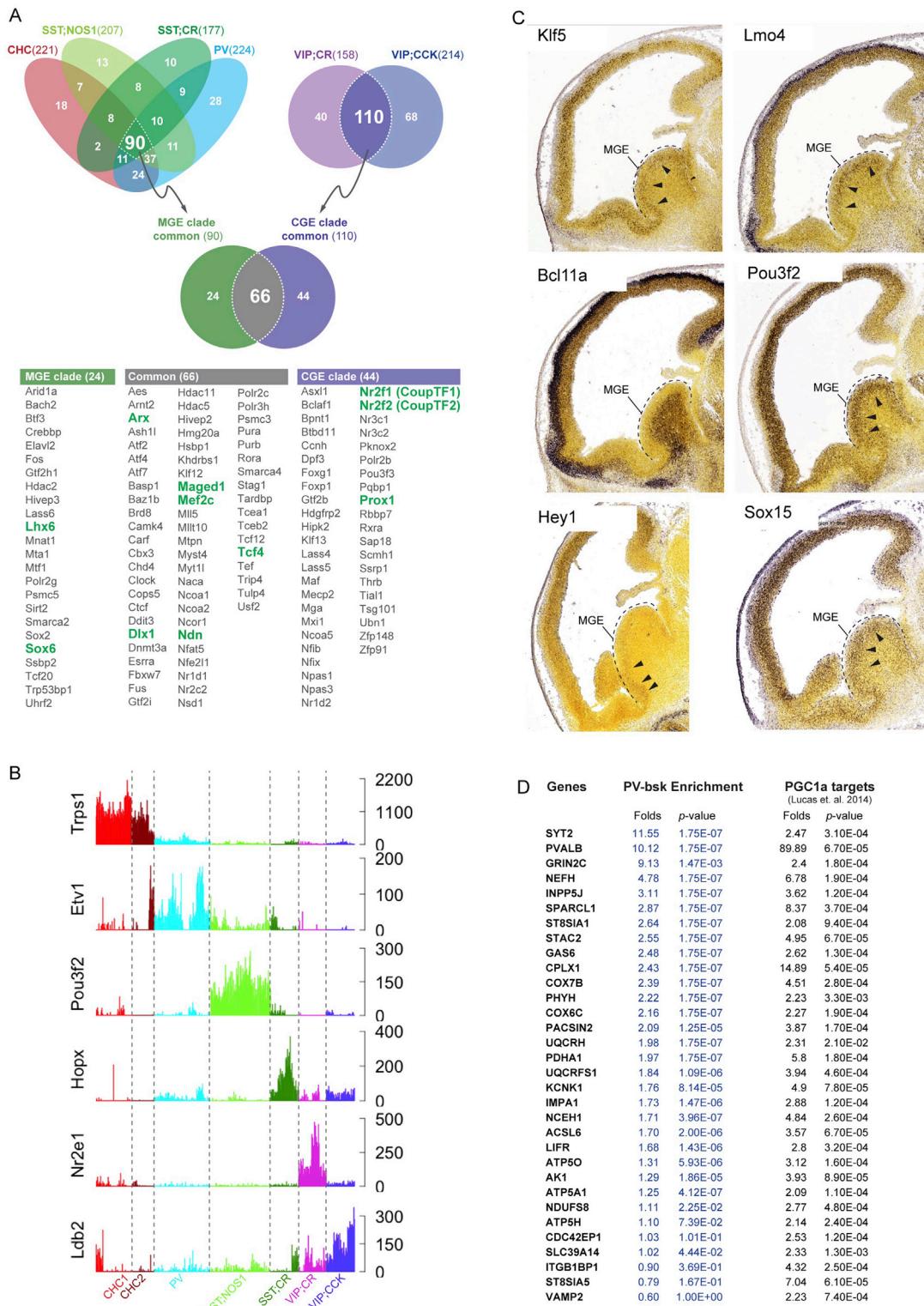


(legend on next page)

---

**Figure S6. SST/NOS1 Axons Are Myelinated, Related to Figure 6**

- (A) Barplot of neuropeptides and endogenous ligands expressed in ON/OFF pattern in individual cells of each PCP (colors).
- (B) Left: Schematic role of PTN in promoting differentiation of oligodendrocyte precursors for axonal myelination. PTN expression in LPCs predicts myelination of their axons. Right: Confirmation of myelination of LPC axons. 3D rendering of a deep layer LPC axon (red) with paranodal CASPR expression (green).
- (C) Genetic knock-in strategy to generate Ptn-CreER mouse line. Inset shows Southern blot of the targeted locus with band at expected ~6.5Kb; clone C4 was used for blastocyst injection. Clones were also confirmed by PCR of 3' and 5' arms (data not shown).
- (D and E) As predicted from Figures 6D and S6A, Ptn labels LPCs. Coronal sections of Ptn-CreER;Ai14 brain upon sparse (low-dose) Tamoxifen induction at P14 labels Sst positive LPCs in L2/3 (C) and L6 (D). P14 induction in Ptn-CreER driver also labeled some pyramidal neurons (arrowheads); this is expected as Ptn is known to be expressed in pyramidal neurons (Fame et al., 2017). CC = corpus callosum, st = striatum.
- (F) Layer 6 Ptn-CreER;Ai14 LPC is also immunopositive for Nos1; arrow shows long axon entering corpus callosum (CC) (He et al., 2016).
- (G) Left: 63X super-resolution image shows SST/NOS1 axons traversing deep cortical layers, among a field of myelin nodes labeled by CASPR immunostaining. Dotted box is enlarged in the right 3 panels, which show 3D rendering of higher magnification images and coaxially apposed CASPR signal over tdTomato-labeled axons of SST/NOS1 at three angular rotations (white arrows).

**Figure S7. Transcription Factor Profiles among PCPs and the PGC1 $\alpha$  Transcription Network, Related to Figure 7**

(A) Top: Venn diagram of all TFs expressed in each PCP; TFs were assigned to a set if it is expressed in  $\geq 75\%$  of cells within a PCP at a level of  $\geq 30\text{uTPM}$  in single cells. Common TFs in MGE and CGE derived PCPs are further intersected to generate a list of common GABAergic TFs. Bottom: Table of TFs in each set, green fonts indicate TFs known to be expressed as computed from the Venn diagram.

(legend continued on next page)

A comparison between MGE versus CGE neurons revealed that ~65 TFs are common between the two populations with ~90 TFs enriched in MGE and ~110 TFs enriched in CGE populations. Among the 4 MGE-derived and 2 CGE-derived PCPs, each expresses between 150-220 TFs ([Figure S7](#)).  
(B) Barplots of TFs expressed in ON/OFF pattern in individual cells of PCPs.  
(C) mRNA *in situ* Images from Allen Developmental Mouse Brain Atlas at E13.5 show expression of transcripts in the MGE that are expressed in PV < SST and SST/CR and SST/NOS1 from [Figures 7F](#) and [7G](#).  
(D) Table showing that > 75% of PGC1 $\alpha$  targets identified by [Lucas et al. \(2014\)](#) are also significantly ( $p < 1 \times 10^{-5}$ ) enriched in PVC transcriptome data.