Connect-seq to superimpose molecular on anatomical neural circuit maps

Naresh K. Hanchate, Eun Jeong Lee, Andria Ellis, Kunio Kondoh, Donghui Kuang, Ryan Basom, Cole Trapnell, and Linda B. Buck

The mouse brain contains about 75 million neurons interconnected in a vast array of neural circuits. The identities and functions of individual neuronal components of most circuits are undefined. Here we describe a method, termed “Connect-seq,” which combines retrograde viral tracing and single-cell transcriptomics to uncover the molecular identities of upstream neurons in a specific circuit and the signaling molecules they use to communicate. Connect-seq can generate a molecular map that can be superimposed on a neuroanatomical map to permit molecular and genetic interrogation of how the neuronal components of a circuit control its function. Application of this method to hypothalamic neurons controlling stress hormone responses to predator stressors, including predator odors, tissue injury, inflammation, and emotional and psychological stress. Studies using classical neuroanatomical and neurophysiological approaches have provided a large body of information on PVN inputs that may affect CRHNs and the effects of classical neurotransmitters on CRHN activity and function. However, the exact identities of upstream neurons that control CRHN functions and the mechanisms by which they do so have not been defined.

Viral tracing studies indicate that CRHNs receive direct synaptic input from neurons in 31 different brain areas. These include 19 areas of the hypothalamus, a brain area that integrates information from the body and external environment and organizes fitting behavioral and physiological responses. Neurons two or more synapses upstream of CRHNs are seen in additional areas, including one small area of the olfactory cortex that has proved key to stress hormone responses to predator odors. The viral tracing studies provide an anatomical map of neurons directly presynaptic to CRHNs, but the molecular identities of those neurons are unknown.

If one could identify genes whose expression defines subsets of neurons upstream of CRHNs in different brain areas, one would have molecular tools to explore which subsets mediate responses to different stressors and to manipulate those subsets to gain insight into how they function. Information regarding neurotransmitters and neuropeptidemodulators used by the upstream neurons to communicate with CRHNs would, in addition, provide information relevant not only to an understanding of neural circuit functions, but also potential insights into pharmacological interventions to modify the functions of specific circuit components.

To test the Connect-seq strategy, we infected Cre recombinase-expressing CRHNs with a Cre-dependent pseudorabies virus (PRV) that travels retrogradely across synapses. Using flow cytometry, we isolated single virus-infected neurons and then used single-cell RNA sequencing (RNA-seq) to define the transcriptomes of individual upstream neurons. These experiments revealed a

Significance

Single-cell transcriptomics has emerged as a powerful means to define the molecular heterogeneity of brain neurons. However, which of the neurons with known transcriptomes interact with each other in specific neural circuits is largely unknown. Here, we devised a strategy, termed “Connect-seq,” which combines retrograde viral tracing and single-cell transcriptomics to determine the molecular identities of individual upstream neurons in a defined circuit. Using Connect-seq, we uncovered a large variety of signaling molecules expressed in neurons upstream of hypothalamic neurons that control physiological responses to stress. Information obtained by Connect-seq can be used to overlay molecular maps on anatomical neural circuit maps and generate molecular tools for probing the functions of individual circuit components.

Data deposition: Raw sequencing data related to this study have been archived in the National Center for Biotechnology Information Gene Expression Omnibus (GEO) database https://www.ncbi.nlm.nih.gov/geo (accession no. GSE139923).


Reviewers: L.L., Stanford University; and H.Z., Allen Institute for Brain Science.

The authors declare no competing interest.

Published under the PNAS license.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.1912176117/-/DCSupplemental.
large diversity of neurotransmitter and neuromodulator signaling molecules in neurons directly upstream of CRHNs, including more than 40 different neuropeptides. Many individual neurons coexpressed multiple different signaling molecules. These included neurons coexpressing neuropeptides with neurotransmitters or biogenic amines, as well as neurons coexpressing different neuropeptides. Upstream neurons expressing specific signaling molecules mapped to selected brain areas, demonstrating that Connect-seq can provide molecular tools with which to dissect the functions of individual neuronal components of neural circuits.

Results

Transcriptome Analysis of Neurons Upstream of CRHNs. We first developed a Pseudorabies virus, PRVB180, which has Cre recombinase-dependent expression of thymidine kinase (TK) fused to green fluorescent protein (GFP) (Fig. 1). After infecting neurons expressing Cre, this virus will travel retrogradely across synapses to infect upstream neurons. We injected PRVB180 into the PVN of CRH-IRES-Cre (CRH-Cre) mice (Fig. 1B). Following infection with PRVB180, immunostaining of brain sections for GFP indicated that the virus travels retrogradely at the same rate as PRVB177, which expresses TK-HA (hemagglutinin-tagged TK) instead of TK-GFP and travels to directly presynaptic neurons on day 3 postinfection (d3pi) and then to neurons two synapses upstream on d4pi (8).

We next conducted transcriptome analysis of neurons upstream of CRHNs. We infected CRHNs with PRVB180 and, on d3pi, isolated the hypothalamus and dissociated the tissue into a single-cell suspension (Fig. 1C). Using flow cytometry, we isolated GFP+ (PRV-infected) cells, one per well, in 96-well plates (14) (Materials and Methods and SI Appendix, Fig. S1). Since they were isolated on d3pi, the neurons isolated should be presynaptic to CRHNs, but the possibility that a minor fraction are two synapses upstream cannot be absolutely excluded. Microscopic examination of sorted cells indicated that all but one or two per well (Materials and Methods and SI Appendix, Fig. S2). cDNAs were prepared from each sorted cell and tested for GFP expression by PCR.

We then conducted RNA-seq (15, 16) on individual GFP-expressing cells by Illumina sequencing cDNAs prepared from those cells (SI Appendix, Fig. S3A). Gene expression in individual cells (SI Appendix, Fig. S3B) was determined by alignment of mouse genome using standard methods (17, 18). Infection of cells with PRV was confirmed by aligning sequence to PRV genome data (Materials and Methods and SI Appendix, Fig. S3C and D). We then examined the transcriptome of each cell for expression of markers of neurons or glia (Materials and Methods) and proceeded to analyze 117 cells that expressed neuronal markers.

Neurons were sequenced at an average of ~6.5 million reads per cell. Comparison of sequenced reads and numbers of genes expressed in neurons did not indicate a correlation (Pearson’s coefficient = −0.06; SI Appendix, Fig. S3E). To evaluate the effect of PRV infection on the number of genes detected, we used Pearson’s correlation. Comparison of the number of genes expressed with the percentage of PRV transcripts in individual neurons indicated a weak correlation (~0.22), suggesting that PRV infection can reduce the number of genes detected to a small extent (SI Appendix, Fig. S3F).

Neurons Upstream of CRHNs Express Diverse Neurotransmitters and Neuromodulators. To characterize the neurons upstream of CRHNs, we focused on signaling molecules that neurons use to convey information to their downstream partners (19). These
include classical “fast neurotransmitters,” which act via ligand-gated ion channels on downstream neurons to rapidly activate or inhibit those neurons, and numerous neuromodulators that bind to G protein-coupled receptors (GPCRs) on downstream neurons to modulate their excitability. We reasoned that the large number of different signaling molecules in the brain and their differential expression among neurons could serve to 1) optimize the discovery of molecular identifiers that would distinguish different upstream neurons and 2) provide potential insight into the molecular mechanisms used by upstream neurons to communicate with CRHNs.

We identified a large number and variety of neurotransmitters and neuromodulators expressed by PRV-infected neurons upstream of CRHNs (Fig. 2). We used a conservative threshold of 1 FPKM (fragments per kilobase of transcript per million mapped reads) to define expressed genes in individual cells. All of the infected neurons that were analyzed expressed at least one signaling molecule. The signaling molecules included glutamate and GABA, which act via ligand-gated ion channels and are the major excitatory and inhibitory neurotransmitters in the brain, respectively. We also identified two other fast neurotransmitters that signal through ligand-gated ion channels, acetylcholine and glycine. All of these neurotransmitters can also act as neuromodulators by binding to specific GPCRs on downstream neurons. The fast neurotransmitters were expressed in 65% of upstream neurons analyzed (Fig. 24), but the number of neurons expressing different neurotransmitters varied (Fig. 2B and SI Appendix, Fig. S4A). Glutamate and GABA were expressed in 24.8% and 36.8% of the neurons, respectively, while the other neurotransmitters were expressed in far fewer neurons (1.7% for acetylcholine and 17.1% for glycine). CRHNs have ligand-gated channels for both glutamate (20) and GABA (21, 22) and receive direct synaptic input from both neurotransmitters (11, 23, 24). Our results suggest that CRHNs can be activated by numerous biogenic amines, but that many other upstream neurons may release GABA and glutamate to CRHNs. Our observations are consistent with a complex pattern of stimulatory, inhibitory, and neuromodulatory effects that CRHNs can exert on whether they are released onto CRHNs together with glutamate or GABA in other neurons (33-35). In our experiments, of seven neurons expressing Cck (cholecystokinin), two coexpressed glutamate and one coexpressed GABA, and, of 17 neurons expressing Tac1 (tachykinin1), three coexpressed glutamate, eight coexpressed GABA, and one coexpressed both. We also found Cck and Tac1 coexpressed with amines in different neurons in RNA-seq data from 898 hypothalamic neurons, which were available in the Gene Expression Omnibus database (accession no. GSE74672) (5).

It is known that neurons expressing a neuropeptide or biogenic amine can also express a classical neurotransmitter (4, 5, 32, 36, 37). These results further indicate that many neuropeptides can be coexpressed with glutamate in some upstream neurons but with GABA in others. Of 39 neuropeptides coexpressed with glutamate or GABA in our studies, 17 were coexpressed with glutamate in some neurons and GABA in others (Fig. 3C, and, interestingly, among these, nine were coexpressed with glutamate and GABA in the same neurons. These results suggest that individual neuropeptides could express different modulatory effects on CRHNs depending on how they are released onto CRHNs together with glutamate versus GABA. For example, if a neuropeptide acts to enhance CRHN excitability, it may further enhance glutamate stimulation of CRHNs, but dampen GABAergic suppression of CRHNs. Our observations are consistent with a complex pattern of stimulatory, inhibitory, and neuromodulatory effects that act at the level of synaptic input to CRHNs from individual neurons to fine tune the effects of upstream inputs on the CRHNs.

Upstream Neurons Can Coexpress Multiple Neuropeptides. These studies also revealed many upstream neurons expressing more than one neuropeptide. As noted earlier, neuropeptides were expressed in 88.9% of the neurons upstream of CRHNs. While 27.4% of upstream neurons expressed a single neuropeptide, 61.5% coexpressed two or more, 6.8% coexpressed five or six neuropeptides, and a few neurons expressed seven or eight neuropeptides each (Fig. 4 A and B). Since single-cell RNA-seq is subject to the loss (“drop-out”) of low copy number mRNAs that can vary among cells, coexpression of different neuropeptides in additional neurons cannot be excluded.

These results are consistent with previous reports that neurons expressing a given neuropeptide can also express other neuropeptides (35, 38, 39), although the large number of neuropeptides...
that could be coexpressed in single neurons was unexpected. Nonetheless, we also found individual neurons expressing multiple neuropeptides in available transcriptome data for 898 single hypothalamic neurons in a previous study (as detailed earlier) (5). In addition, we compared RNA-seq data on PRV+\textit{Pomc}+ neurons isolated by Connect-seq versus uninfected \textit{Pomc}+ neurons manually isolated from \textit{Pomc}\text{-eGFP} mice (SI Appendix, Fig. S6). In both cases, individual \textit{Pomc}+ neurons coexpressed one or more other neuropeptides with \textit{Pomc}.

Comparisons of different neuropeptides coexpressed in individual neurons indicate that a given neuropeptide can be coexpressed with a variety of other neuropeptides (Fig. 4C). In some cases, this was only one other neuropeptide, but, in others, a neuropeptide was coexpressed with numerous other neuropeptides, although often in different neurons. The level of expression of individual neuropeptides and other signaling molecules varied among neurons. However, high-level expression of two neuropeptides in the same neuron was relatively rare. For

Fig. 2. Neurons upstream of CRHNs express a large array of signaling molecules. (A) Pie charts show percentages of upstream neurons expressing different neurotransmitters, biogenic amines, neuropeptides, gaseous neuromodulators, and other neuromodulators. (B) Heat map illustrating expression levels of different signaling molecules expressed in 117 individual upstream neurons. Different neurons are indicated on the x-axis. The y-axis shows different signaling molecules, neuronal markers, and housekeeping genes and their encoding or indicator genes.
Penk (proenkephalin) was coexpressed with 12 other neuropeptides, but high levels (≥1,000 FPKM) of only two of those neuropeptides were detected in a neuron(s) expressing a high level of Penk (Fig. 4C).

To further investigate the coexpression of neuropeptides in single neurons, we compared the levels of expression of different neuropeptides in individual neurons (Fig. 5). At the level of single neurons, a few neurons expressed two or three neuropeptides at similar levels. However, many other neurons expressed one neuropeptide at a much higher level than one or more others expressed in the same cell. These results suggest that, while a neuron might express multiple neuropeptides, the downstream effects of one might predominate over the effects of others. Some neuropeptides, as well as other signaling molecules, were expressed at very different levels in different cells (Fig. 5 and SI Appendix, Fig. S7), raising the possibility that they might play a more important role in signaling to CRHNs by some upstream neurons than others. One intriguing question is whether the relative expression levels of different signaling molecules in a neuron can be altered according to physiological state, as previously reported for one neuropeptide (40).

**Upstream Neurons Expressing Specific Neuromodulators Map to Selected Brain Areas.** The major goal of these studies was to find a way to uncover molecular identifiers of upstream neurons that would allow for a molecular map to be superimposed on the anatomical map of neurons upstream of CRHNs. To investigate this possibility, we examined neurons upstream of CRHNs for the expression of neuromodulators we had identified by transcriptome analysis of upstream neurons.

In our studies, we identified numerous neuropeptides in upstream neurons isolated from the hypothalamus. Viral tracing studies previously showed neurons directly upstream of CRHNs in 19 hypothalamic areas. Using in situ hybridization data available in the Allen Brain Atlas (http://mouse.brain-map.org/) (41), we confirmed that a number of the neuropeptides we had identified in upstream neurons are indeed expressed in one or more of those 19 hypothalamic areas (SI Appendix, Fig. S8). Of example, Penk (proenkephalin) was coexpressed with 12 other neuropeptides, but high levels (≥1,000 FPKM) of only two of those neuropeptides were detected in a neuron(s) expressing a high level of Penk (Fig. 4C).
course, it was unclear whether neurons seen in the atlas are upstream of CRHNs.

To investigate this question, we analyzed the locations of upstream neurons expressing specific neuromodulators. We first infected Cre-expressing CRHNs with PRVB177, which has Cre-dependent expression of hemagglutinin (HA)-tagged TK (8). After 3 days, when the virus had crossed one synapse to infect immediately upstream neurons, we costained brain sections with riboprobes for specific neuromodulators and anti-HA antibodies to detect PRV-infected neurons expressing detectable levels of neuromodulators (Fig. 5).

For detecting PRV in the brain sections, we used a commercial viral detection kit. After 3 days, we observed the expression of HA-tagged TK in the infected neurons. The selected expression of neuromodulators that we had identified in CRHNS compared to the numbers used in large-scale transcriptome census studies of the brain. Even so, the upstream neurons expressed a large number and variety of neurotransmitters and neuromodulators, including dozens of different neuropeptides. While some neurons expressed only one of these signaling molecules, others expressed different combinations. Since the upstream neurons could synapse with many other downstream neurons in addition to CRHNS, it is possible that only some of these signaling molecules are used to communicate with CRHNS. Nonetheless, these findings suggest the potential for a complex array of inputs to CRHNS that could excite or inhibit CRHNS or modulate their excitability.

Localization experiments revealed subsets of upstream neurons expressing individual neuromodulators we had identified by Connect-seq in distinct brain areas. In conclusion, Connect-seq enables the construction of a molecular map that can be superimposed on an anatomical map of neural circuits upstream of CRHNS. In the present studies, we devised a strategy, termed Connect-seq, which utilizes a combination of viral tracing and single-cell transcriptomics to recover the molecular identities of individual neurons upstream of a specific set of neurons and reveal the neurotransmitters and neuromodulators they use to communicate with their downstream synaptic targets. As a proof of concept, we used Connect-seq to gain insight into circuits that transmit information to CRHNS, which control blood levels of stress hormones.

We examined a relatively small number of neurons upstream of CRHNS compared to the numbers used in large-scale transcriptome census studies of the brain. Even so, the upstream neurons expressed a large number and variety of neurotransmitters and neuromodulators, including dozens of different neuropeptides. While some neurons expressed only one of these signaling molecules, others expressed different combinations. Since the upstream neurons could synapse with many other downstream neurons in addition to CRHNS, it is possible that only some of these signaling molecules are used to communicate with CRHNS. Nonetheless, these findings suggest the potential for a complex array of inputs to CRHNS that could excite or inhibit CRHNS or modulate their excitability.

Localization experiments revealed subsets of upstream neurons expressing individual neuromodulators we had identified by Connect-seq in distinct brain areas. In conclusion, Connect-seq enables the construction of a molecular map that can be superimposed on an anatomical map of neural circuits, thereby allowing the investigation of roles played by individual neuronal components of those circuits under normal conditions and in disease.

Materials and Methods

Mice. CRH-IRES-Cre (CRH-Cre) mice were generated as described previously (13). All procedures involving mice were approved by the Fred Hutchinson Cancer Research Center Institutional Animal Care and Use Committee. Pomc-eGFP (stock no. 009593) and C57BL/6J mice were purchased from The Jackson Laboratory.
Pseudorabies Viral Vectors. Construction of PRVB177 was described previously (8). PRVB180 was made according to the methods described previously (8). Briefly, PRVB180 was constructed using homologous recombination between a targeting vector (PRVTK-GFP) and genomic DNA of PRV TK-BaBlu, a thymidine kinase (TK)-deleted PRV Bartha strain derivative with a LacZ insertion into the gG locus (50). For generating PRVTK-GFP, a flexstop-flanked sequence (51) encoding a PRV TK fused at its C terminus to enhanced green fluorescent protein (eGFP), obtained from the pEGFP-N1 vector (Clontech), was first cloned with an inverse orientation into an eGFP-deleted pEGFP-N1 vector (Clontech). Next, following restriction digestion, NsiI fragments containing a CMV promoter, the flexstop-flanked coding sequence, and an SV40 polyadenylation signal were cloned between gG locus sequences matching those 5’ and 3’ to the lacZ sequence in PRV TK-BaBlu to give the final targeting vector (PRVTK-GFP). The vector was then linearized and cotransfected with PRV TK-BaBlu genomic DNA into HEK 293T cells (ATCC). Recombinant virus clones were selected and confirmed following methods described previously (52).

PRVs were propagated by infecting PK15 cells (ATCC) with the viruses using a multiplicity of infection (m.o.i.) of 0.1 to 0.01. After ∼2 days of infection, cells showed a prominent cytopathic effect. Cells were scraped from the dishes, pelleted by centrifugation, and frozen using liquid nitrogen and then quickly thawed in a 37 °C water bath. After three freeze–thaw cycles, cell debris was removed by centrifugation twice at 1,000 × g for 5 min, and the supernatant was aliquoted and stored at −80 °C until use. The titer of viral stocks was determined using standard plaque assays on PK15 cells (53), with titers expressed in plaque-forming units (p.f.u.).

Stereotaxic Injections. Viruses were injected into the PVN of CRH-Cre mice as described previously (8). PRVB180 was made according to the methods described previously (8). Briefly, PRVB180 was constructed using homologous recombination between a targeting vector (PRVTK-GFP) and genomic DNA of PRV TK-BaBlu, a thymidine kinase (TK)-deleted PRV Bartha strain derivative with a LacZ insertion into the gG locus (50). For generating PRVTK-GFP, a flexstop-flanked sequence (51) encoding a PRV TK fused at its C terminus to enhanced green fluorescent protein (eGFP), obtained from the pEGFP-N1 vector (Clontech), was first cloned with an inverse orientation into an eGFP-deleted pEGFP-N1 vector (Clontech). Next, following restriction digestion, NsiI fragments containing a CMV promoter, the flexstop-flanked coding sequence, and an SV40 polyadenylation signal were cloned between gG locus sequences matching those 5’ and 3’ to the lacZ sequence in PRV TK-BaBlu to give the final targeting vector (PRVTK-GFP). The vector was then linearized and cotransfected with PRV TK-BaBlu genomic DNA into HEK 293T cells (ATCC). Recombinant virus clones were selected and confirmed following methods described previously (52).

PRVs were propagated by infecting PK15 cells (ATCC) with the viruses using a multiplicity of infection (m.o.i.) of 0.1 to 0.01. After ∼2 days of infection, cells showed a prominent cytopathic effect. Cells were scraped from the dishes, pelleted by centrifugation, and frozen using liquid nitrogen and then quickly thawed in a 37 °C water bath. After three freeze–thaw cycles, cell debris was removed by centrifugation twice at 1,000 × g for 5 min, and the supernatant was aliquoted and stored at −80 °C until use. The titer of viral stocks was determined using standard plaque assays on PK15 cells (53), with titers expressed in plaque-forming units (p.f.u.).

Stereotaxic Injections. Viruses were injected into the PVN of CRH-Cre mice as described previously (8). All injections were done under inhalation anesthesia of 2% isoflurane. Briefly, 1 µL of PRVs (PRVB180, PRVB177, 1 to 1.5 × 10^6 p.f.u.) were loaded into a 1-µL syringe and injected bilaterally into the

---

**Fig. 6.** Upstream neurons expressing individual signaling molecules map to specific brain areas. (A) Graphs show the percentage of PRV + neurons labeled for different neuropeptides (Avp, Tac1, Npy, or Hdc) in different brain areas following CRHN infection with PRVB177. Error bars indicate SEM. Parenthesized numbers indicate the number of animals (“n”) per condition. (B) CRHNs were infected with PRVB177 and brain sections costained on d3pi with neuropeptide riboprobes (green) and anti-HA antibodies (red; PRV + cells). Arrows indicate colabeled neurons. (Scale bars, 25 µM.)
brain at a rate of 100 nl/min using a Stereotaxic Alignment System (David Kopf Instruments). The needle was inserted into the PBN using a stereotaxic atlas (54). After recovery, animals were singly housed with regular 12-h dark/light cycles, and food and water were provided ad libitum.

**Isolation of Single Cells.** For Connect-seq experiments, we used a total of 34 adult (6 to 16 wk old) virgin male and female mice for 34 independent experiments. Eleven adults (five males and six females) yielded 698 GFP+ cells, and 384 cells were sequenced. For female mice, the day of estrous cycle was not determined. Adult CRH-Cre mice were injected with PRV180. After 3 days, mice were euthanized by cervical dislocation, and the brain quickly removed and submerged in ice-cold Hibernate-A medium (A1247501; Thermo Fisher Scientific). The hypothalamus was carefully microdissected under a microscope by obtaining a single coronal section extending from the optic chiasm to the posterior end of the hypothalamus (rostral-caudal axis) and then the hypothalamus between the anterior commissures (laterally and dorsally), which can contain ventral portions of the BNST. The isolated tissue was dissected into tiny pieces in dissociation buffer [Hibernate-A medium with papain (10 U mL−1; PAP; PDS kit; Worthington Biochemical) and DNase (200 U mL−1; DNase vial, D2; PDS kit; Worthington Biochemical)]. Dissected tissue fragments were transferred into 5 mL of dissociation buffer and incubated for 30 min at room temperature. After incubation, tissue pieces were gently triturated 2 to 3 times through a series of fire-polished glass Pasteur pipettes with decreasing diameters of ~600 μm, ~300 μm, and ~150 μm to dissociate tissue into a cloudy suspension. Cells were sieved using a 70-μm Pasteur pipette with decreasing diameters of 300, 150, and 100 μm and then analyzed to sort GFP+ cells with high GFP and low DAPI (to obtain live cells) were inventoried first running the “cuffquant” tool on the aligned reads for each cell with the “−u” option, which performs additional algorithmic steps designed to better assign ambiguously mapped reads to the correct gene of origin. Per-cell gene-expression profiles were subsequently normalized with the “cuffnorm” utility, using the “classic-fpmk” normalization method, for use in downstream analysis.

A total of 384 PRV-infected cells were sequenced at a depth average of 6,674,000 reads (median, ~6,700,000; range, ~17,500 to ~13,300,000). A total of 347 cells that expressed at least 500 genes per cell were considered for downstream analysis. These included 117 neurons (as detailed later), which were sequenced at an average depth of 6,530,693 reads (median, 6,696,084; range, 34,414 to 12,211,210) and expressed an average of 3,272 genes (median, 3,215; range, 548 to 10,391).

Expression of genes encoding signaling molecules in single-cell transcriptome data were analyzed using a binary operator “%in%” in R (58, 59). Briefly, lists of gene names of different signaling molecules and the transcriptome data on upstream neurons were uploaded into R and converted to a vector and a data frame, respectively. The data frames with expression data contained gene names in rows and different cells arranged in columns. Using the binary operator function, first, the column in the expression data that contained gene names and the vector with gene symbols in signaling molecules were analyzed to find a match, and, as a result, a new data frame was generated that contained the expression data of signaling molecules in individual upstream neurons. Similarly, to analyze the gene expression of signaling molecules in previous sequencing data, we used the binary operator function in R. Previous sequencing data were obtained from Gene Expression Omnibus (accession no. GSE74672) (5), which contained expression data of 898 neurons, as defined by their cluster algorithm. Marker genes that expressed at least one FPKM or at least one molecule (for UMI data) were considered for further analysis. Plots for illustration were generated in RStudio (59) using ggplot2 package (60).
Analysis of Marker Genes. Previous studies established that PRVs spread faithfully between functionally connected neurons by direct cell-to-cell contact, but not through follower cells or transcellular routes (61, 62). Neuronal infection with PRVs can activate neighboring astrocytes and microglia, which isolate severely infected and dying neurons. However, glial cells lack the machinery required for the production and release of infectious (enveloped) virions (63, 64).

We first manually inspected transcriptome data for transcripts of different cell type markers in order to include neurons, but exclude nonneuronal cells. We used a stringent threshold of 10 FPKM for cell type markers as in our previous analysis of olfactory sensory neurons (55). Cells were classified using cell type markers in order to include neurons, but exclude nonneuronal cells.

Cells were classified using a criterion that a cell must express two or more of six different marker genes of a cell type but not more than one of any other cell type. We used the previous analysis of olfactory sensory neurons (55). Cells were classified using a threshold of 0.05% Tween buffer and then treated using the TSA-plus FLU kit (Perkin-Elmer). Sections were then incubated with 0.5 μg mL−1 DAPI and Alexa555-Streptavidin (1:1,000, no. 32355; Thermo Fisher) at room temperature for 1 h and then coverslipped with Fluoromount-G (no. 0100-01; Southern Biotech).

Cell Counting. Sections were analyzed essentially as described previously (8). Briefly, images were collected using an AxioCam camera and AxioZoom.Z2 microscope with an apodeme device (Zeiss). Images were acquired using the auto-exposure setting because of the variability in background signals between sections in different animals. No additional postprocessing was performed on collected images for counting. Counting was performed blindly. A mouse brain atlas was used to identify brain structures microscopically and in digital photos. Every fifth section was analyzed for all experiments. Brain areas containing colabeled cells for a given signaling molecule (at least 10 labeled cells in a given area) in at least two animals were included.

Abbreviations for Brain Areas. Abbreviations used for brain areas are according to a mouse brain atlas (54) and our previous report (8): AH, anterior hypothalamic area; ARAC, arcuate hypothalamic nucleus; AVPe, anterointerventricular nucleus; BNSTa, bed nucleus of the stria terminalis, anterior part; BNSTp, bed nucleus of the stria terminalis, posterior part; DMH, dorsomedial hypothalamic nucleus; LH, lateral hypothalamic area; LPAG, lateral preoptic area; LGP, lateral paragigantocellular nucleus; LPO, lateral preoptic area; LS, lateral septal nucleus; MnPO, median preoptic nucleus; MPO, medial preoptic nucleus; MTu, medial tuberal nucleus; NTS, nucleus of the solitary tract; PBN, parabrachial nucleus; Pe, periventricular nucleus of the hypothalamus; PH, posterior hypothalamic nucleus; PLH, peduncular part of lateral hypothalamic area; PMV, premammillary nucleus; ventral part; SC, suprachiasmatic nucleus; SHy, septohypothalamic nucleus; VMM, ventromedial hypothalamic nucleus; and ZI, zona incerta.

ACKNOWLEDGMENTS. We thank J. Delrow, A. Marty, A. Dawson, and R. Meredith at the Fred Hutchinson Cancer Research Center (FHRC) Genomics Facility for their assistance with RNA-seq and A. Berger, S. Dozono, and B. Raden, at the FHRC Flow Cytometry Facility for their assistance with flow sorting. We thank X. Ye, A. Spray, and E. Allbrecht for technical assistance and members of the Buck Laboratory for helpful discussions. This work was supported by the Millen Literary Trust (E.J.L.), the Howard Hughes Medical Institute (L.B.B.), NIH Grants R01 DC013052 (L.B.B.), R01 DC014642 (L.B.B.), and EY008158 (C.T.), the Paul G. Allen Frontiers Foundation (through the Allen Discovery Center for Cell Lineage Tracking), and an Alfred P. Sloan Fellowship (C.T.). L.B.B. is on the Board of Directors of International Flavors & Fragrances.

56. bcft2fastq conversion software (version 1.8.4) (Illumina, Inc., San Diego, CA).
60. H. Wickham, ggplot2: Elegant graphics for data analysis (Springer-Verlag, New York, NY, 2016).
Supplementary information for

Connect-seq to superimpose molecular on anatomical neural circuit maps

Naresh K. Hanchate\textsuperscript{a,1}, Eun Jeong Lee\textsuperscript{a,1}, Andria Ellis\textsuperscript{b,2}, Kunio Kondoh\textsuperscript{a,c,2}, Donghui Kuang\textsuperscript{a}, Ryan Basom\textsuperscript{d}, Cole Trapnell\textsuperscript{b,e}, and Linda B. Buck\textsuperscript{a}

\textsuperscript{1}N.K.H. and E.J.L. contributed equally to this work.
\textsuperscript{2}A.E. and K.K. contributed equally to this work.

Corresponding author: L.B.B. (email: lbuck@fhcrc.org, phone: 206-667-6316)

This PDF file includes: Figs. S1 to S9
Fig. S1. Workflow of flow cytometry. Dissociated cells from cell suspensions were sorted using FACS in a “single cell sorting mode”. A representative of FACS illustrating different gates to place individual GFP+ DAPI- (live PRV+) cells singly into wells of 96 well plates. Gate 1 was applied to exclude debris and sort cells based on size and granularity using forward and side scatter area pulse parameters, Gates 2 and 3 were applied to exclude doublets using pulse height (H) and width (W) parameters, and Gate 4 was applied to sort cells based on their fluorescence. Live cell-impermeable DAPI was used to stain and exclude dead cells. On average, 2.8±0.6% of cells were found within the DAPI+ gate settings. Single cells in the boxed GFP+ area were isolated in 96 well plates.
Fig. S2. Quality control to assess the use of flow cytometry to sort one cell per well in multiwell plates. CRHNs were infected with PRVB180 and dissociated cells from the hypothalamus subjected to flow cytometry. Individual GFP+/DAPI- cells were sorted one per well into wells of 60-well Terasaki plates and microscopy used to examine individual wells for fluorescent and non-fluorescent cells. A total of 125 wells contained a single GFP+ cell. None contained more than one cell. Shown are representative low magnification images of individual wells with insets showing higher magnification bright-field (white background) and fluorescent images (black background) of the cell found in each well. Arrows indicate the location of the cell observed in each well.
Fig. S3. Quality metrics for single cell RNA-seq libraries. A-D. Graphs show the number of sequenced reads per cell (A), number of genes detected at or above one FPKM (B), percentage of reads mapping to exons of mouse genome (GENCODE M15, mouse genome build mm10) (grey) or PRV genome (NC_006151.1) (blue) (C), and number of reads mapping to the PRV genome (D). Individual cells are indicated by bars and colored according to cell type, as indicated. E-F. Scatterplots show comparison of the number of genes detected in neurons and sequencing depth (sequenced reads) (E) or level of PRV infection (percentage of PRV transcripts) (F). Each dot indicates a single neuron. Blue lines indicate linear regression of correlation.
Supplementary figure 4

Fig. S4. Variations in proportions of upstream neurons expressing different signaling molecules. Graphs show the number of upstream neurons expressing different neurotransmitters (A), biogenic amines (B), other neuromodulators (C), and neuropeptides (D).
Fig. S5. Coexpression of different signaling molecules in upstream neurons. Graphs show the number of upstream neurons that coexpressed biogenic amines (A) or other neuromodulators (B) with different neurotransmitters. Graphs show the number of upstream neurons that coexpressed individual neuropeptides with gaseous neuromodulators (C) or other neuromodulators (D).
Fig. S6. Comparison of neuropeptides expressed in Pomc+ neurons isolated by Connect-seq or manual isolation. (A) Graph shows the average (colored bars) number of genes detected in individual Pomc+ cells (black dots) isolated manually from Pomc-eGFP mice or by using Connect-seq. Error bars indicate s.e.m. (B) Heatmaps illustrate diverse coexpression of neuropeptides in Pomc+ cells. Neuropeptides are indicated on the y-axis and individual cell identifier numbers on the x-axis. Blue boxes indicate neuropeptides coexpressed in Pomc+ neurons obtained using both methods.
Supplementary figure 7

A

Neuropeptide

B

Signaling molecule (marker gene)

Fig. S7. Expression levels of signaling molecules vary in upstream neurons

Graphs show the expression levels of different neuropeptides (A) and marker genes for other signaling molecules (B) in individual neurons (blue dots) upstream of CRHNs. Data are shown as log-transformed FPKM.
Supplementary figure 8

Fig. S8. Expression of genes encoding neuromodulators in different areas of the hypothalamus.

Chart shows data obtained from the Allen Brain Atlas in situ hybridization database indicating expression of specific neuropeptides or histamine in areas of the hypothalamus found to contain neurons upstream of CRHNs (green boxes) in previous viral tracing studies (8).
**Fig. S9. Connect-seq for superimposing molecular on anatomical circuit maps.** Retrograde viral tracing provided an anatomical map of neurons upstream of CRHNs. Connect-seq defined the transcriptomes of single neurons upstream of CRHNs and revealed signaling molecules they express. By mapping the locations of upstream neurons expressing those signaling molecules, it is possible to superimpose a molecular map on the anatomical map of neural circuits upstream of CRHNs.