HHS Public Access

A repeated molecular architecture across thalamic pathways

James W. Phillips[1,3,6,*], **Anton Schulmann**[1,2,6], **Erina Hara**[1,4], **Johan Winnubst**[1], **Chenghao Liu**[2], **Vera Valakh**[2], **Lihua Wang**[1], **Brenda C. Shields**[1,5], **Wyatt Korff**[1], **Jayaram Chandrashekar**[1], **Andrew L. Lemire**[1], **Brett Mensh**[1], **Joshua T. Dudman**[1], **Sacha B. Nelson**[1,2,*], **Adam W. Hantman**[1,*]

[1]HHMI Janelia Research Campus, Ashburn, VA, USA

[2]Brandeis University, Waltham, MA, USA

[3]Department of Physiology, Development and Neuroscience, University of Cambridge, UK

[4]Present Address: Institute for Genomic Medicine, Columbia University Medical Center, New York, NY

[5]Present Address: Duke University - Pratt School of Engineering, Durham, NC

[6]These authors contributed equally (co-first authorship): James W. Phillips, Anton Schulmann

Abstract

Thalamus is the central communication hub of the forebrain, providing cerebral cortex with inputs from sensory organs, subcortical systems, and cortex itself. Multiple thalamic regions send convergent information to each cortical region, but the organizational logic of thalamic projections has remained elusive. Through comprehensive transcriptional analyses of retrogradely labeled thalamic neurons in adult mice, we identify three major profiles of thalamic pathway. These profiles exist along a continuum that is repeated across all major projection systems, such as those for vision, motor control, and cognition. The largest component of gene expression variation in mouse thalamus is topographically organized with features conserved in humans. Transcriptional differences between these thalamic neuronal identities are tied to cellular features critical for function, such as axonal morphology and membrane properties. Molecular profiling therefore reveals covariation in properties of thalamic pathways serving all major input modalities and output targets, establishing a molecular framework for understanding thalamus.

*Correspondence to: James Phillips (jameswphillips91@gmail.com), Sacha Nelson (nelson@brandeis.edu), Adam Hantman (hantmana@janelia.hhmi.org).

INTRODUCTION

The wide-ranging connectivity of the thalamus makes it the central communication hub of the forebrain. Its pathways carry almost all peripheral and subcortical information that reaches the neocortex, and it also facilitates information transfer between cortical areas[1–3]. The functions of the forebrain, and especially the neocortex, thus build on the properties of thalamic projection neurons. Given that cell-type specialization underlies the functioning of all understood organs, knowing whether there are multiple types of thalamic projection neuron – and their properties – is likely a necessary step for dissecting the logic of how thalamo-cortical projections support cortical function[4].

Thalamus is divided into approximately thirty distinct nuclei, based mainly upon cytoarchitectural and connectomic features[2]. There have been multiple, largely complementary approaches for classifying thalamic nuclei. The oldest approach, dating at least to Thomas Willis, groups nuclei by the type of sensory input that they receive, distinguishing visual, auditory, and somatosensory nuclei[2,5]. This was extended to non-sensory regions by considering the output targets of each nucleus, thus defining motor, prefrontal, anterior (limbic), and hippocampal-related projection systems[2]. However, this classification does not explain why multiple thalamic nuclei exist within each projection system.

Different thalamic nuclei projecting to a given cortical area can vary substantially in features relevant for understanding neural computation[4], and several theories use these features to categorize thalamic nuclei. One theory divides thalamic nuclei into those that receive their main 'driving' input from subcortical or peripheral regions ('first order'), and those that receive an input from cortical layer 5 ('higher order')[1]. In this view, these higher order nuclei serve to facilitate cortico-cortical communication. A second approach, based primarily upon lesion and stimulation studies, distinguishes between the arousal-related intralaminar nuclei and the rest of thalamus, with these nuclei being especially critical to maintaining signatures of alertness[6,7]. Finally, the core/matrix theory divides the thalamus into 'core' nuclei that project in a narrow, topographic manner to the middle layers, and 'matrix' nuclei that project over a broad area of layer 1 of neocortex[8–10]. The core nuclei, which overlap those defined as 'first order,' are argued to 'form the basis for sensory and other inputs that are to be used as a basis for perception', while the matrix nuclei are argued to underlie the 'binding of multiple aspects of sensory experience'[9]. These theories suggest that there are common types of thalamic projection neuron repeated across projection systems, implying shared functional specializations.

There are, however, limitations to these theories. First, there is significant variation in the functional and anatomical properties of nuclei within each proposed class. For example, the 'matrix' class from the core/matrix theory also includes the intralaminar nuclei, which have significantly different morphology and function than other matrix nuclei[7,11,12]. Second, the core/matrix and first/higher order theories are almost exclusively defined and tested in a sensory context. Whether they apply to the various projection systems subserving navigation, emotion, and motor control, for example, is largely unknown. Third, the categorizations are based only on a single anatomical feature, while neurons vary over many

parameters. Which of these schemes accounts for the largest amount of variation in neuronal identities therefore remains unclear.

In other brain regions, molecular analysis has provided a principled way to characterize neural populations, and resulting cell-type classifications have provided frameworks for functional understanding[13]. For example, broad molecular interneuron classes are delineated across neocortex and hippocampus by molecular markers such as parvalbumin (*Pvalb*) or somatostatin (*Sst*), and these cell classes have been linked to differing roles in brain function, such as oscillations and participation in disinhibitory circuits[14].

Our understanding of the cellular architecture of thalamus at the molecular level lags that of other brain regions and is largely limited to *in situ* hybridization or microarray experiments from either a limited set of molecular markers, a limited set of thalamic nuclei, or both[15–17]. Though in one theory, core/matrix, expression of *Pvalb* and calbindin (*Calb1*) mark a proposed dichotomy, these molecular markers were identified fortuitously rather than in an unsupervised manner from a large-scale screen[8]. It is therefore not established that this division actually marks the major branch of cellular heterogeneity in thalamic pathways, meaning there could be other principles of thalamic molecular variation that better link neuronal identity to function.

Here we use dense transcriptional profiling to map the dominant features organizing thalamic projection neurons. We find that almost all projection systems are composed of three broad major molecular profiles, which lie along a spectrum of relatively continuous cellular variation. Molecular differences between thalamic cell profiles are tied to anatomical and electrophysiological properties important for neural function. Together, our results both describe the molecular variation underlying thalamic cell type and reveal similarities in the properties of thalamic pathways serving diverse cognitive systems. This establishes a common, molecularly defined framework for understanding the central communication hub of the forebrain.

RESULTS

A near-comprehensive transcriptomic atlas of thalamic pathways

To understand the organization of thalamic pathways, we combined anatomical and genetic approaches to produce a near-comprehensive, projection-specific, transcriptomic atlas of murine thalamus at the nuclei level (http://thalamoseq.janelia.org). Thalamic nuclei were retrogradely labeled from forebrain areas, such as visual, somatosensory, and motor cortices. The separate nuclei were then microdissected via comparison with the Paxinos Atlas[18], cells pooled, and RNA isolated and sequenced (RNAseq for 8 projection targets, 22 nuclei, 120 samples; for information on labelling for each nucleus see Supplementary Tables 1 and 2). This pooled-cell approach gains greater sensitivity for detecting low-expressing genes such as receptors and ion channels, at the cost of averaging over cells[19]. Anterograde tracing of inputs to thalamus was used to guide microdissections when identification of nuclear boundaries was ambiguous (Fig. 1a,b and Supplementary Table 1, which contains information on each dissection approach). Several approaches were used to assess the data

quality and accuracy of dissections, including comparison of samples from the same thalamic nucleus obtained via different labelling approaches (see Supplementary Fig. 1).

Taking advantage of this near-comprehensive resource, we performed hierarchical clustering to explore the major branches of thalamic molecular heterogeneity (on the 500 most differentially expressed genes, see methods, Supplementary Fig. 1c and Source Data 1). We identified five major subdivisions of nuclei across thalamus (Fig. 1c). Anterior dorsal nucleus (AD) and nucleus reuniens (RE) each formed profiles of their own. This left three major multi-nuclei thalamic gene expression profiles. These major profiles were not explained by cortical projection target, since the multiple nuclei projecting to motor, somatosensory or visual cortices split across the three different profiles (Supplementary Fig. 2). Thus, each cortical region examined receives input from across these three genetically defined profiles.

Nuclei comprising these molecular profiles lie in a topographic arrangement. Such spatial arrangement is a feature of existing theories, with laterally-localized 'first-order' sensory nuclei, such as the dorsal lateral geniculate nucleus (LGd) and ventrobasal nuclear complex (VB), included in one profile, 'higher-order' nuclei, such as the lateral posterior nucleus (LP) and the posterior nuclear complex (PO), included in the middle profiles, and midline and intralaminar nuclei included in our most medial profiles (Fig. 1d). Though this overlaps with existing theories, it also suggests substantial heterogeneity within the known matrix class from the core/matrix classification[8–10]. Matrix nuclei, marked by *Calb1* expression, are split by the first branch of hierarchical clustering (see Supplementary Fig. 2 for additional analysis). By revealing shared molecular profiles in thalamus, our analysis of molecular heterogeneity recapitulates and extends existing notions of a cross-projection organizational scheme.

Thalamic gene expression variation is low-dimensional, topographic, and likely functionally relevant

Identifying the patterns of gene expression variation underlying these thalamic profiles could hold important clues for understanding their origin and function. To do this we used principal component analysis (PCA) to find the major sources of variation in our high-dimensional gene expression data (again using the top 500 differentially expressed genes). This revealed 6 significant components (Fig. 2a, Supplementary Fig. 3). As with our hierarchical clustering, the main division was not drawn by projection classes, whose contribution to the variance across nuclei was relatively small (Fig. 2a) and which mainly served to separate primary sensory and motor cortical targets from other regions (Supplementary Fig. 4 and 5). Instead, the largest principal component separated nuclei by the profiles from hierarchical clustering. Despite the large number of nuclei analyzed, more than half (53.3%) of the variation was accounted for by the first two principal components, with the leading dimension of variation alone accounting for 38.3% of the total variance (Fig. 2b, Supplementary Fig. 3a for plot of variance explained). Thalamic transcriptional variation is thus concentrated in a small number of dimensions.

Based on their relative order on this first component, we named the three major profiles primary, secondary, and tertiary. This naming reflects three features. Firstly, nuclei

conveying relatively unprocessed, 'first-order' sensory information are in our primary group, while intralaminar nuclei, linked to more cognitive signals such as arousal, are included in our tertiary group (Fig. 1d, Fig. 2b). Secondly, the three profiles are spatially organized in a sequential topographic arrangement, with the leading axis (PC1) selectively correlating with mediolateral position (Fig. 2c, Supplementary Fig. 6). Thirdly, a progressive expression difference exists from primary through to tertiary nuclei, evident in the number of genes differentially expressed between the groups, with the primary and tertiary nuclei being most separate, and the other two comparisons being less so (Fig. 2d). Thus the major linear axis of transcriptional diversity in thalamus encompasses existing functional groups and reflects a progressive, topographic shift across thalamus.

Examination of this largest transcriptional axis revealed additional, within-profile heterogeneity. For example, calretinin (*Calb2*) labeled only the leftward-most (in Fig. 2b) tertiary nuclei from PC1, while the other tertiary nuclei were not substantially marked (Supplementary Fig. 2). These nuclei are also split by a lower branch of the hierarchical clustering (Fig. 1c). This suggests that our major profiles describe positions along a leading axis of variance, rather than classic, relatively homogenous 'cell types' defined by unique marker genes.

We next examined whether this leading axis of variation was present in humans. To do this, we projected previously published post-mortem human thalamus microarray data[20] onto the same axis using the mouse gene loadings, finding that it produced a similar separation of the nuclei (Fig. 2e). As in the mouse data, the LGd mapped onto the positive end (p=0.031; one-sided Wilcoxon signed-rank test) and the intralaminar nuclei on the negative end (p=0.023 for rostral and caudal intralaminar nuclei, ILr and ILc, respectively; one-sided Wilcoxon signed-rank test; Fig. 2e) of PC1, demonstrating at least some conservation of this axis between mice and humans.

The gene classes most strongly underlying this leading axis of variance were diverse, including those encoding proteins involved in synaptic vesicle release, metabolic enzymes, kinases, and calcium binding proteins (Fig. 2f). However, most enriched were genes encoding neurotransmitter receptors, ion channels, and signaling molecules (Fig. 2g). This suggests that this axis of variance is linked to functional differences.

Variation of functionally relevant genes and electrophysiological properties across nuclear profiles

Given the prominent role of receptor and ion channel expression on the principle axis of variance, we sought to assess whether these profiles indeed differed in properties related to neuronal function. We performed PCA on the expression profiles of voltage-gated ion channel or neurotransmitter/modulator receptor encoding genes (Fig. 3a, left and right respectively). Analysis with these limited gene sets reproduced a similar separation of profiles in PC1 (Fig. 3a), confirming that ion channel and receptor profiles are organized along the same axis identified in Fig. 2. This also revealed a relative enrichment of neuromodulator receptors toward the tertiary end of the axis (Fig. 3b, Supplementary Fig. 7), while genes linked to high firing rates via faster channel kinetics, such as K_v3 channels (*Kcnc1*, *Kcnc3*), the $Na_v1.6$ channel (*Scn8a*), and the $K_v\beta3$ subunit (*Kcnab3*)[21–24], tended to

be progressively elevated toward primary profile nuclei (Fig. 3b, Supplementary Fig. 7). This raised the possibility that action potential width may progressively narrow from tertiary to primary nuclei, a metric more easily predicted from gene expression than other neuronal properties.

Whole-cell recordings from the motor-related nuclei CM, VA, and VL (representing the three main nuclear profiles; Fig. 1c) confirmed a variation of physiological features between nuclei (Fig. 3c,d). Neurons recorded within VL have the narrowest action potential width and those in CM have the widest. In addition, many other electrophysiological properties showed a systematic gradient ranging from VL through VA to CM, including synaptic properties (Fig. 3c,d and Supplementary Fig. 8). Prior work has shown substantial differences in electrophysiological properties between different thalamic nuclei, but to date this has not been incorporated into thalamic organizational schemes[25–27]. Here, we link electrophysiological properties with the three major transcriptional profiles of thalamic neuron.

Variation of axonal morphology across nuclear profiles—The relationship between gene expression variation and electrophysiological features led us to ask if a relationship also existed with morphological features. To answer this, we combined anatomical characterization of individual neurons with inferred gene expression profiles. We reconstructed the axo-dendritic morphology of 106 sparsely-labeled thalamic neurons over the entire brain and aligned the resulting reconstructions to the 3D reference space of the Allen Mouse Brain Atlas (Fig. 4)[28]. Example neurons resulting from this are shown in Fig. 4a. Alignment to the reference volume allowed use of the Allen Institute's In Situ Hybridization Atlas data to infer the likely gene expression of neurons reconstructed from each region of thalamus. For this, a subset of genes were chosen that strongly mark the leading axis of variance identified in this paper, and the voxel-wise expression of these genes was reduced to a single dimension using PCA (genes shown in Fig. 4b along with inferred expression in each neuron). Each reconstructed neuron therefore had an inferred PC1 score (Fig. 4c,d). As in Fig. 2a and 3a, a positive PC1 score corresponded to a primary-like neuron, while a negative PC1 score corresponded to a tertiary-like neuron.

The reconstructed neurons are shown in Fig. 4e, with each line representing a single neuron, ranked by their inferred position on PC1. This revealed that axonal targeting of specific cortical layers was correlated to PC1 position (Fig. 4e,f). For example, neurons situated in PC1 positions occupied by the primary profile had increased axon length in the middle layers of cerebral cortex (Fig. 4f,g). Contrastingly, secondary profile neurons had extensive layer 1 arborizations (Fig. 4f,h), corresponding to the classic 'matrix'-like cells. The layer 1 arborization was least prominent in tertiary neurons, despite their expression of matrix marker *Calb1*, supporting the distinction made by our molecular analysis. Layer 1 innervation by the primary group was in between secondary and tertiary, consistent with prior reports of a layer 1 arborization in a subset of primary sensory thalamic neurons[29].

Even within a given profile, neurons could exhibit quite substantial morphological heterogeneity. For example, while some neurons in tertiary nuclei had very large caudoputamen projections, others were much smaller (Fig. 4i). Within-profile variation was

also found for other features such as innervation of specific cortical layers and total axon density in cortex (Fig. 4g,h, Supplementary Fig. 9). Previous classification schemes have often used morphological properties to define discrete thalamic types[1,9,30]. However, our results suggest a relatively continuous, systematic variation of axonal morphology along the largest, topographic axis of gene expression variation.

Single-cell heterogeneity within each projection system

Thus far we analyzed thalamic gene expression profiles across pools of cells from different nuclei from various projection systems. While this allows comparisons across nuclei, it does not allow assessment of within-nuclei heterogeneity. To probe heterogeneity within projection systems at greater resolution, we profiled the transcriptomes of single thalamic neurons projecting to motor, somatosensory, and visual cortex (Supplementary Table 3; for quality control see Supplementary Fig. 10). In addition, the single-cell approach allowed us to assess auditory and prefrontal projection systems, which we were not able to subdissect for pooled-cell approaches. We note that our single-cell collection was biased against tertiary nuclei due to issues related to sparsity (see Fig. 5 legend for more detail).

Analysis of single-cell RNAseq data from all projection systems together revealed many clusters (Fig. 5a, visualized by t-distributed stochastic neighbor embedding; tSNE). Most individual clusters were a mixture of different projection systems (Fig. 5b,c), with only one cluster being specific for a single projection target (Fig. 5c, cluster 9). All projections were split over at least 5 clusters, but several clusters were enriched for specific projections. This likely reflects a combined influence of projection system and profile on neuronal identity; projection identity also explained smaller aspects of the variance of the pooled-cell datasets (Fig. 2a, Supplementary Fig. 4 and 5). In addition to confirming the existence of thalamic cell types shared across projection systems, the main axis of variance seen in our pooled-cell RNAseq dataset was recapitulated in the single-cell dataset (Fig. 5d). Thalamic neurons were distributed along the first principal component derived from our pooled-cell RNAseq dataset, as shown in Fig. 5d, in which each single cell is colored by its position on the pooled-cell PC1. This was also true when each projection system was examined separately (Fig. 5e). The distribution along PC1 appeared to be relatively continuous, rather than discrete. Notably, a relationship to PC1 was also found for auditory and prefrontal projection neurons, which were not assessed in pooled-cell RNAseq experiments (Fig. 5e).

We next sought to identify the major cell types within each projection system, irrespective of nuclear boundaries (Fig. 6). We therefore clustered each projection system separately, revealing that each projection was characterized by multiple clusters (Fig. 6a). Similar genes marked clusters for each projection system, suggesting a shared molecular architecture (Fig. 6b). When plotted along the PC1 from the pooled-cell data, the clusters each occupied a different, characteristic position, consistent with them being separated by a similar axis of variation to the largest variance component from the pooled-cell data (Fig. 6c, Supplementary Fig. 11). Differences along the pooled-cell PC1 are therefore present across projection systems at the single-cell level. For some projection systems, multiple clusters could be mapped within the range of a single profile identified in the pooled-cell experiments. For example, though *Tnnt1* and *Pvalb* are both enriched toward the primary

end of PC1, they appear in separate single-cell clusters for neurons projecting to motor cortex. Multiple clusters related to a single profile provide evidence of further heterogeneity not appreciated in the pooled-cell RNAseq experiments.

The single-cell RNAseq datasets indicated relatively continuous variation between cell clusters of a projection system (Fig. 6c, Supplementary Fig. 12a). This suggests the presence of intermediate cells that express markers of more than one cell type. To test this, we examined three genes (*Tnnt1*, *Necab1*, *Calb2*), which marked the clusters of many of the projection systems. Although single-cell RNAseq is inherently noisy, we found evidence for cells expressing more than one of these genes (Supplementary Fig. 12b). To confirm the existence of these intermediate cells and to spatially map them, we performed multiplexed fluorescent in situ hybridization (multi-FISH) for these three single-cell RNAseq cluster marker genes. Neurons expressing *Tnnt1*, *Necab1*, and *Calb2* were largely spatially separated over the mediolateral axis of thalamus (Fig. 7a). This was even true for thalamic areas not profiled by single-cell RNAseq, such as the three navigation-related nuclei (reuniens, rhomboid, and submedial thalamic (SMT) nucleus; Fig. 7a,b). However, clearly labeled intermediate cells were also identified, with most intermediate cells being located at the boundaries between thalamic nuclei (Fig. 7b, Supplementary Fig. 13 and 14). Markers for primary and tertiary showed substantial coexpression in the reuniens/rhomboid/SMT system, a finding at odds with the notion of a perfect continuum in every system.

Intermediate cells were also found within accepted nuclear boundaries. For example, cells expressing either *Pvalb* or *Tnnt1* alone exist in the periphery of VL. However, cells expressing both could be found in the middle of VL (Fig. 7c). These mixtures of cell types within an individual nucleus likely explain why multiple single-cell clusters map to an individual profile. Thus, using a combination of single-cell RNAseq and multi-FISH, we found spatially separated neuronal identities across all thalamic pathways. Intermediate cells bridge the major clusters creating continuous transitions between classes of thalamic neurons.

Discussion

The properties and organization of thalamic pathways have a critical influence upon the rest of the forebrain. They have been studied at multiple levels of analysis, but to date their molecular architecture has remained elusive compared to other major brain regions. Here, through near-comprehensive transcriptional profiling, we identify conserved molecular variation that is repeated across thalamic pathways serving diverse aspects of cognition, an architecture upon which further specializations for each projection system are laid (see Supplementary Fig. 15 for visual summary).

We interpret this central result as supporting theories of a homologous architecture that extends across different thalamic pathways. The broad molecular profiles identified here share substantial commonalities with existing functional divisions from the core/matrix and first/higher order frameworks[1,9,10], suggesting that these theories indeed reflect fundamental principles of thalamic organization at the molecular level. For example, nuclei providing relatively unfiltered sensory information (core/first-order) are in our primary molecular

profile, while layer-5 input recipient nuclei with broad cortical layer 1 arborization (matrix/higher-order) linked to synchronization of large populations of cortical neurons are in our secondary profile[1,9]. Though earlier work had questioned whether theories from primate would readily translate to other species, we find that *Pvalb* (core) and *Calb1* (matrix) mRNA mark different molecular branches in rodent as well, though not the main branch[8]. Variations across species will of course exist, but our analysis of human microarray data suggests cross-species conservation of the basic features revealed here.

Our molecular findings also extend these theories in several ways. Though these theories were largely developed and tested in sensory systems, we find that a shared axis of gene expression variation actually extends across almost all thalamic systems, substantially expanding these theories' likely relevance for understanding thalamocortical communication. Our profiling also identifies substantially greater cellular heterogeneity than was apparent with a small set of molecular markers. This includes the 'matrix' class being composed of at least two quite different molecular identities, consistent with intralaminar/midline (our tertiary) nuclei being distinct from the classic sensory 'higher-order' pathways. Importantly, here we also establish an ordered relationship of increasing molecular difference from primary to secondary to tertiary profiles. Thus, by identifying an organization in a statistically unsupervised manner based upon a broad set of genes, we identify a molecular framework upon which an increasingly detailed understanding of thalamus can be built.

Existing theories emphasize discrete cellular types, however we find intermediate cells between profiles. Coexistence of features can be found in previous data relevant for first- and higher-order distinctions and for core/matrix markers[8,29,31,32]. Our RNAseq data shows that, at least in mouse, neurons co-expressing two markers indeed reflect intermediate cell types at the transcriptome level (Fig. 6, Supplementary Fig. 12). Intermediate cells are localized mostly at the boundary zones between nuclei, however we also found intermediate cell types within individual nuclei (e.g. VL). This suggests that the major axis of molecular variation extends in a topographic (topogenetic) manner across the thalamus, with intermediate cells connecting the major classes.

Differences in functional properties such as neuromodulator responses and burst firing patterns have been shown between thalamic nuclei[33,34]. Here, we find that genes directly determining functional properties are a major component of the molecular variation of thalamic profiles. Whole-cell recordings and full single-cell reconstructions showed covariation of this genetic axis with important neuronal features such as intrinsic electrophysiological signatures, synaptic properties, and axonal projection differences. Narrower and lower-threshold action potentials were found in the primary group, while tertiary nuclei showed prominent neuromodulator receptor expression enrichment, with secondary intermediate. This may hint at the molecular, mechanistic underpinnings for their specialized roles that include accurately relaying peripheral sensory information and modulation of forebrain state respectively[35–37].

Importantly, finding this repeated genetic variation across projection systems does not preclude the existence of other additional axes of variance, and specializations for each

projection system are indeed increasingly appreciated[38]. We also found substantial gene expression differences between projection systems, though marker genes for projection systems were rare (Fig. 6b, Source Data 3). This included substantial differences between motor and visual thalamus, where potentially corresponding anatomical differences have recently also been found. For example, layer 5 inputs from frontal cortex to VM (secondary motor thalamus) are now thought to be at least partly closed-loop, enabling maintenance of persistent activity, while inputs from visual cortex to LP (secondary visual thalamus) are thought to be open-loop, allowing information exchange between different areas[39–41]. Thalamic neuronal identities are thus defined by a combination of projection target and profile information.

Though our data come from adult mice, thalamic gene expression is known to be highly dynamic during development[42]. Prior work has shown a link between a thalamic neuron's birth date and its resulting spatial location and calcium binding protein expression (*Calb1*, *Calb2* and *Pvalb*), including reports of a latero-medial gradient of neuronal birth date in thalamus[43–46]. Speculatively, this suggests a sequential development of the topogenetic axis identified here. This process is likely driven by both extrinsic and intrinsic factors, as specification of thalamic neuronal identity is known to rely upon a reciprocal interaction with its many input targets[16,47,48]. The relatively topographic organization of gene expression variation in rodent thalamus may make it an ideal model for understanding how continua of cellular properties are formed and what function they serve[49].

Here we identify a pattern of transcriptional variation that is repeated across almost all thalamic projection systems and is enriched in neuronal effector genes including ion channels and receptors. These results suggest a foundation for a conserved functional organization of thalamic pathways.

Accession codes

Raw and processed RNAseq data were deposited to the Gene Expression Omnibus (GSE133911 for pooled-cell thalamic nuclei-level RNAseq data and GSE133912 for the single-cell RNAseq data).

Data Availability

The gene expression data can be browsed, plotted and differential gene expression tests can be performed interactively at http://thalamoseq.janelia.org. The single-neuron reconstruction data is available at http://ml-neuronbrowser.janelia.org. Scripts for data analysis are available on github (https://github.com/aschulmann/ThalamoSeq).

Online Methods

EXPERIMENTAL METHODS

Animal care—Experimental procedures were approved by the Institutional Animal Care and Use Committee at the Janelia Research Campus. Mice were housed on a 12-hour light/ dark cycle, with *ad libitum* food and water. All mice were bred against C57Bl/6J background. The majority of mice used for the transcriptomic studies were 8–12 weeks old

(see Source Data 1 and 3 for more details). A list of transgenic mice used in this study can be found in Supplementary Table 5.

Acquisition of samples—Cells were fluorescently labeled to enable manual dissection. This was done through retrograde labelling via either viral or tracer injection into the major projection field of the nucleus of interest. For viral injections, rAAV2-retro expressing Cre-dependent GFP or tdTomato under the CAG promoter were injected, with volumes of 50–100 nL at each depth[51]. Minimum survival time for viral injections was 3 weeks post-injection. Viruses were prepared by Janelia Virus Services. Non-viral retrograde tracer labelling used the lipophilic tracer DiI (2.5mg/mL in DMSO, injecting volumes of 50–200 nL per site, from Molecular probes) or Lumafluor red retrobeads (diluted 3x in PBS, 50–200 nL per site). Minimum survival times for non-viral retrograde tracing was 4 days post-injection. Anterograde labelling of inputs to thalamus was also used in a small number of cases (see Supplementary Tables 1 and 2). Injection coordinates are described in Supplementary Tables 2 and 3.

We referred to the Paxinos and Franklin mouse brain atlas to guide our dissections[18]. For the majority of thalamic regions, retrograde tracers labeled populations corresponding to identified thalamic nuclei (Supplementary Tables 1 and 2 for targeting details). However, the caudal intralaminar nuclei (parafascicular complex) were less clearly delineated. This likely reflects additional heterogeneity within this complex beyond that shown in atlases[52,53].

Data collection and analysis were not performed blind to the conditions of the experiments. Data were not perfectly randomized in collection, but an effort was made to distribute biological replicates across batches. Source Data 1 and 3 include all the relevant sample metadata.

Manual cell sorting and RNAseq

Sorted pooled-cell RNAseq: Fluorescent cells were collected and sequenced as previously described[19,54,55]. Briefly, animals were deeply anaesthetized with isoflurane and euthanized. Coronal slices (200–300 μm) were cut and placed for 1 hour at room temperature with pronases and neural activity blockers in artificial cerebrospinal fluid (ACSF). Relevant regions were then microdissected, and the tissue dissociated. The resulting cell suspensions were diluted with filtered ACSF and washed at least 3 times by transferring them to clean dishes. This process produces negligible contamination with non-fluorescent tissue (Supplementary Fig. 1a)[19]. After the final wash, samples were aspirated in a small volume (3 μl) and lysed in 47 μl XB lysis buffer using the Picopure kit (KIT0204, Thermo Fisher Scientific) in a 200 μL PCR tube (Axygen), incubated for 5 min at 42°C on a thermal cycler and stored at −80°C.

Library preparation and sequencing was performed by the Janelia Quantitative Genomics core. RNA was isolated from each sample using the PicoPure RNA isolation kit (Life Technologies) with on-column RNAase-free DNase I treatment (Qiagen). 1 μL External RNA Controls Consortium (ERCC) RNA spike-in mix at 10^{-5} dilution (Life Technologies) was added to each sample. Amplification was then performed using the Ovation RNA-seq v2 kit (NuGEN), yielding 4–8 μg of cDNA. cDNA was sheared using Covaris E220

sonicator according to the manufacturer's recommendations, and the Ovation rapid DR multiplexing kit (NuGEN) was used to make libraries for sequencing. Libraries were quantified by qPCR using Kapa Library Quantification for Illumina (Kapa Biosystems KK4854). Equimolar pools were made and sequenced on a HiSeq 2500 (Illumina).

Sorted single-cell RNAseq: Samples for single-cell RNAseq were prepared as described previously[56]. Retrogradely labeled cells were isolated as described above, and collected separately into 8-well strips containing 3 μL Smart-Seq2 lysis buffer, flash-frozen on dry ice, and stored at −80°C until further use[57].

Upon thawing, cells were lysed and digested with Proteinase K, and 1 μL ERCC RNA spike-in mix at 10^{-7} dilution (Life Technologies) and barcoded RT primers were added. cDNA was synthesized using the Maxima H Minus Reverse Transcriptase (Thermo Fisher Scientific) and E5V6NEXT template switch oligo, followed by heat inactivation of reverse transcriptase. PCR amplification using the HiFi PCR kit (Kapa Biosystems) and SINGV6 primer was performed with a modified thermocycling protocol (98°C for 3 min, 20 cycles of 98°C for 20 s, 64°C for 15 s, 72°C for 4 min, final extension at 72°C for 5 min). Samples were then pooled across strips, purified with Ampure XP beads (Beckman Coulter), washed twice with 70% ethanol and eluted in water. These pooled strips were then combined to create the plate-level cDNA pool for tagmentation, and concentration was determined using Qubit High-Sensitivity DNA kit (Thermo Fisher Scientific).

Tagmentation and library preparation using 600 pg cDNA from each plate of cells was then performed with a modified Nextera XT (Illumina) protocol, but using the P5NEXTPT5 primer and tagmentation time extended to 15 minutes[58]. The libraries were then purified following the Nextera XT protocol (at 0.6x ratio) and quantified by qPCR using Kapa Library Quantification (Kapa Biosystems). 6–10 plates were run on a NextSeq 550 high-output flow cell. Read 1 contained the cell barcode and unique molecular identifier (UMI). Read 2 contained a cDNA fragment from the 3' end of the transcript.

Multi-FISH—C57Bl/6J mice (~8 weeks old) were anesthetized with isoflurane then fixed via transcardial perfusion with PBS followed by 4% paraformaldehyde in PBS, pH 7.4. Brains were post-fixed at 4°C overnight, washed 3 times with PBS, and cryoprotected in a sucrose series of 10%, 20% then 30% in PBS at 4°C. All solutions were prepared RNase-free. Brains were sectioned (14 μm) on a Leica CM3050S cryostat, mounted onto Fisher SuperFrost Plus slides, and stored at −80°C.

Multi-FISH was performed using the RNAscope Multiplex Fluorescent Assay platform from ACDBio, following the manufacturer's protocol. The probes used were: *Calb2* (ref 313641), *Necab1* (ref 428541), *Tnnt1* (ref 466911), and *Pvalb* (ref 421931). Fluorescent dyes were DAPI, Alexa Fluor 488, Atto 550, and Atto 647. Images were acquired using a Zeiss LSM 880 confocal microscope, with an air 20x (0.8 Numerical Aperture) objective unless otherwise specified.

For display purposes, each image channel was independently brightened and background subtracted using Zen software (version 2.3 SP1, Zeiss). Gamma correction was applied for

the retrograde beads image in Fig. 1b. All changes were applied uniformly across the entire image.

Electrophysiology—Acute brain slices were prepared from P20–25 mice (mixed male and female). Animals were deeply anesthetized with ketamine/xylazine/acepromizine and transcardially perfused with ice-cold oxygenated cutting solution containing (in mM): 74 NaCl, 3 KCl, 1 NaH2PO$_4$, 25 NaHCO$_3$, 6 MgCl$_2$, 0.5 CaCl$_2$, 5 Sodium Ascorbate, 75 Sucrose, 10 Glucose. 300 μm coronal slices containing the thalamus were cut on a vibratome (Leica), and then recovered for 15 min at 33°C and for 15 min at room temperature in oxygenated cutting solution followed by at least another 1 hour at room temperature in oxygenated ACSF containing (in mM): 126 NaCl, 3 KCl, 1 NaH2PO$_4$, 25 NaHCO$_3$, 2 MgCl$_2$, 2 CaCl$_2$, 10 Glucose. During recordings, slices were perfused with oxygenated 34–35°C ASCF. For current clamp recordings of intrinsic properties, ACSF included 35 μM d,l-2-amino-5-phosphonovaleric acid (APV), 20 μM 6,7-dinitroquinoxaline-2,3-dione (DNQX) to block glutamatergic synaptic transmission and 50 μM picrotoxin to block GABAergic synaptic transmission. For voltage clamp recordings of synaptic events, ACSF included TTX (0.5 μM), APV as above, and DNQX (for mIPSCs) or picrotoxin (for mEPSCs). Target neurons in CM, VA, and VL were identified based on their distance to the mammillothalamic tract and nuclear borders were confirmed with calbindin-D28K (*Calb1*) immunostaining *post hoc*. For the intrinsic properties and mEPSC recordings whole-cell recording pipettes (6 – 8 MΩ) were filled with internal solution containing (in mM): 100 K-gluconate, 20 KCl, 10 HEPES, 4 Mg-ATP, 0.3 Na-GTP, 10 Na-phosphocreatine, and 0.1% biocytin and for mIPSCs containing: 120 KCl, 10 HEPES, 2 MgSO$_4$, 4 Mg-ATP, 0.3 Na-GTP, 10 Na-phosphocreatine, and 0.1% biocytin. Recordings were obtained with MultiClamp 700B amplifiers (Molecular Devices) digitized at 10 kHz using IGOR Pro (WaveMetrics). Resting membrane potentials were adjusted to −70 mV and steady state series resistance was compensated. Series and input resistances were calculated from −5 mV (voltage clamp) or −25 pA (current clamp) steps before each trial of recording. Current clamp liquid junction potentials (measured as −10 mV) were compensated *post hoc*[24]. Neurons with high series resistance (> 25 MΩ current clamp; >20 MΩ Voltage clamp) or membrane potentials that changed by > 3 mV were excluded. Action potential (AP) properties were measured in current clamp during a series of positive current steps. For each current clamp recorded neuron, the AP threshold, amplitude, afterhyperpolarization (AHP) and width at half-height are average values from the 16th-19th action potentials in trials with 20 to 40 Hz firing rates. Threshold was defined as the membrane potential at which the slope exceeded 10 V/s. Amplitude was the difference between the AP peak and threshold values. The medium AHP was measured as membrane potential hyperpolarization after each action potential relative to threshold. The slow AHP was measured as the hyperpolarization following the trains of 20–40 Hz APs evoked by positive current steps, relative to baseline membrane potential preceding the step. Sag ratio was defined as the ratio of the steady-state hyperpolarization at the end of a 3s −100 pA current step to the maximum hyperpolarization. Spontaneous miniature PSCs were recorded in voltage clamp holding the membrane potential at −70 mV. Custom IGOR scripts and python programs were used to analyze the data. The decay time constant, frequency and amplitude were quantified from mEPSCs and mIPSCs. The decay time constant was calculated by fitting the decay phase of the signal to

an exponential function. The frequency was the number of the mPSCs over the recording time and the amplitude is the average across all mPSCs in each cell.

Immunohistochemistry—After recordings, slices were fixed with 4% paraformaldehyde and 2.5% sucrose in 5x phosphate-buffered saline (PBS) at 4°C for 2–10 days. After washing with PBS, slices were blocked in PBS with 0.3% Triton and 3% BSA at 4°C overnight and then incubated in PBS with 0.3% Triton and 3% BSA and rabbit anti-calbindin D-28k (Swant, CB38a, 1:1000) at 4°C overnight. After washing, they were incubated in PBS with 0.3% Triton, 3% BSA and 5% goat serum with fluorescent protein conjugated goat anti-rabbit IgG (Thermo Fisher Scientific, A11034, 1:1000) and streptavidin (Invitrogen, 1:1000) at 4°C overnight to label *Calb1*-expressing neurons and biocytin-filled neurons.

Complete axonal reconstructions of individual thalamic neurons—Single cell reconstructions were gathered from sparsely labeled brain samples using methods described previously[28]. In short, adult female C57Bl/6 mice were injected with a mixture of viruses consisting of 1:30,000 diluted AAV 2/5 Syn-Cre mixed with 10 parts of high titer ($>10^{12}$ GC/mL) AAV 2/5 CAG-Flex-tdTomato or CAG-Flex-eGFP. Different parts of the thalamus were targeted with the following injection locations (A/P, M/L (depth)): −0.95, −1.0 (−3.3, −2.8); −1.0, −1.2 (−3.7, −3.3); +1.4, −0.8 (−4.0, −3.5); −2.25, −1.5 (−2.8); injection volume was 20nL per location. After allowing for 6–8 weeks of fluorescent protein expression the animals were fixed by transcardial perfusion with 4% paraformaldehyde. Collected brains were then cleared and imaged in their entirety in roughly 7 days using an automated serial two-photon tomography system[28]. Each neuron was traced by two independent tracers, using the Janelia Workstation software package after which a joint consensus tracing was generated[28,59]. Downsampled version of the imaged brains (~5×5×15μm voxel size) were registered to the averaged Allen reference template (10×10×10μm voxel size) using a landmark-based registration approach (3D Slicer; https://www.slicer.org). First, an automated intensity based affine registration was performed to align the sample to the reference atlas (BRAINSFit module[60]). Next, an iterative landmark-registration process, using thin-plate splines approximation, was used to achieve a more precise registration (~150 μm variation). The result of the registration was used to morph traced neurons into the Allen Institute Common Coordinate Framework. In this way neurons from different samples could be pooled together for further analysis. All reconstructions are available online in the MouseLight Neuron Browser (mouselight.janelia.org). Neurons were linked to genetic expression data using the Allen ISH database[61]. Gene expression energy scores at the soma location of each neuron were looked up for the following genes and experiment numbers: *Necab1* (#724727910), *Calb2* (#79556662), *Calb1* (#79556672), *Snca* (#79908848), *Slc17a7* (#79556738), *Pvalb* (#79556738), *Tnnt1* (#77280576), *Kcnab3* (#73817925). Normalized cortical depth information was taken from the Allen Institute Common Coordinate Framework. In short, Laplace's equations were solved between the pia and white matter surfaces to generate normalized equi-potential surfaces that could be used as an analog for cortical depth. Cutoff values for classifying middle layers in the cortex were determined using the range of values enclosed by all layer 4 and 5 cortical structures (0.40–0.78). For layer 1, the range of the cortical depth was 0.92–1. Volume and density calculations were

performed by dividing the cortex into 200 μm isotropic voxels, only voxels containing over 150 μm worth of axons were considered. Positional error was estimated by registering the same brain sample twice and these reconstructions had a difference of 110 ± 60 μm (mean ± standard deviation).

ANALYSIS METHODS

Pooled-cell RNAseq analysis

Data processing and quality control: After removing Illumina adapter sequences using cutadapt, reads were mapped to the mouse reference genome (mm10) using STAR (version 2.5.3a) with 'ENCODE settings' for RNAseq[62]. Mean mapping rate was 82.29% with a standard deviation of 2.25%. One pooled-cell RNAseq sample had a much lower mapping rate compared to all others (<70%) and was excluded. Unique unambiguous exon-mapping reads were summarized at the gene level with featureCounts (version 1.5.2) using GENCODE version M13.

Contamination with common astrocytic, oligodendrocytic, erythrocytic, and microglial transcripts was low, consistent with a lack of substantial contamination by non-fluorescent cells (Supplementary Fig. 1a). Quantification of ERCC spike-in control RNA indicated that our RNAseq profiling was highly accurate and sensitive (Supplementary Fig. 1b). To ensure the specificity of our dissections and to control for potential batch effects, we collected several nuclei through multiple independent labelling approaches, and showed that these samples cluster in a similar manner (Supplementary Fig. 1d).

Differential gene expression: Differential expression was assessed using the Bioconductor package *edgeR*[63]. Low counts were removed by requiring a Transcripts per million (TPM) > 10 in at least 3 samples. This yielded a list of approximately 17,000 expressed genes. Counts were then fitted to a negative binomial generalized linear model, where each factor level represents a different thalamic nucleus, and a likelihood ratio test was used to assess differential expression between groups. p-values were adjusted for multiple tests using the Benjamini-Hochberg method. Genes with false-discovery rate < 0.05 were considered differentially expressed. For selecting the most differentially expressed genes between any thalamic nuclei, we used an ANOVA-like test (ANODEV test for generalized linear models, as described in edgeR User manual 3.2.6), testing for differences between any of the 22 nuclei, and used the 500 genes with the lowest p-value.

To avoid bias due to differences in sample number when comparing numbers of differentially expressed genes between different profiles in Fig. 2c, the groups were subsampled (with replacement) to the size of the smallest group. Bootstrapped log2 fold changes were obtained over 100 iterations. For visualization, clustering, and machine learning of gene expression data, we used variance-stabilized counts produced by the variance-stabilizing transformation in the *DESeq2* R package)[64,65]. All gene expression heatmaps, unless otherwise specified, are Z-scores (mean-centered and scaled by standard deviation) of variance-stabilized gene expression values.

For assessing the role of modality vs. hierarchical class on distinguishing thalamic nuclei, we used elastic-net regularized logistic regression classifiers. Models were trained with different numbers of randomly selected genes as features over 100 iterations. To avoid bias due to variable group size, groups were subsampled to the size of the smallest group. Model tuning was performed using the *glmnet* and *caret* packages in R, and accuracy of the best model was assessed via 5-fold cross-validation.

Unsupervised clustering and functional enrichments: Hierarchical clustering was performed using 1 - Spearman's correlation as a distance metric and complete linkage for agglomeration. Groups were defined by splitting the tree at the level of 5 branches. We termed these profiles, not clusters, as we do not mean to imply discreteness between the classifications. To identify the major split across nuclear profiles in Supplementary Fig. 2, we used three unsupervised clustering approaches: hierarchical clustering as described above cut at the first branch; k-means clustering with k=2 using the *kmeans* function in R; and two-component Gaussian mixture modeling via the *mclust* package in R.

PCA was done using the singular value decomposition based *prcomp* function in R. To assess significance of PCs, we compared the obtained PCs to shuffled data (1000 times randomly permuted order of values for each gene). A PC was considered significant when the actual PC explained at least 50% more variance than the shuffled data more than 95% of the time. This condition was satisfied by the top 6 PCs. We also noted that three sex-related genes (*Xist*, *Eif2s3y*, *Ddx3y*) were among the top 500 differentially expressed genes. *Xist* was removed from the heatmap illustrating the expression of most highly loaded genes in PCA and motor thalamus marker genes as it likely represented a false-positive due to imbalanced sampling of males and females. Inclusion of the three sex-related genes did not have a major effect on the loadings of the other genes. When analyzing the relationship between profiles and projections in Fig. 2a and Supplementary Fig. 4b, only nuclei defined by cortical projection target were used because our striatal injections did not target striatal subregions. For functional enrichment of differentially expressed genes, we used the PANTHER Protein Class Ontology (http://data.pantherdb.org/PANTHER13/ontology/Protein_Class_13.0), which is a consolidated version of molecular function gene ontology. Over-representation in the top 100 genes with the highest PC1 loadings was assessed via two-sided Fisher's exact test. For defining voltage-gated ion channels and neurotransmitter receptors, we downloaded the IUPHAR/BPS database (http://www.guidetopharmacology.org/DATA/targets_and_families.csv). Voltage-gated ion channels were the genes defined as such in the database, while for neurotransmitter receptors we included ionotropic and metabotropic receptors for glutamate, GABA, glycine, acetylcholine, monoamines, neuropeptides, adenosine, and cannabinoids (see Supplementary Fig. 7).

Cross-species comparison: Human brain normalized microarray expression data was downloaded from the Allen Human Brain Atlas (http://human.brain-map.org/static/download)[20]. Expression values for the nine dissected thalamic areas (Supplementary Table 4) were averaged across probes targeting the same gene and across left and right hemispheres when both sides were available. The human microarray data was then quantile-

normalized to the distribution of the grand mean of the mouse RNAseq data. For matching orthologs between mice and human, we used BioMart gene annotation (resulting in 448 one-to-one human matches for the top 500 differentially expressed mouse genes). Human samples were projected onto mouse PC1 by multiplying the quantile-normalized human data (dot product) by the mouse principal component loadings for the matching orthologous genes. Differences of the PC1 scores between the 9 human thalamic areas were assessed using a Kruskal-Wallis test. Significance of PC1 scores of LGd being positive and those of ILr and ILc being negative were assessed via one-sided one-sample Wilcoxon signed-rank test with p-values adjusted for multiple tests using the Benjamini-Hochberg method.

Single-cell RNAseq analysis

Data processing and quality control: Single-cell RNAseq data was trimmed for adapters using cutadapt and aligned to the mouse genome (mm10) using STAR (version 2.5.3a)[62]. To demultiplex cells, collapse UMIs and produce gene-wise counts for each cell, we used a modified version of the *Drop-seq_tools-1.13* pipeline (http://mccarrolllab.com/download/1276/). Briefly, read 1 was tagged based on the cell barcode and UMI, and this information was added to read 2 by merging back the reads after mapping, followed by gene-wise tagging of reads that map onto exons and summarization of digital counts.

Single cells were required to have more than 20,000 UMIs and more than 2,500 genes detected per cell, which yielded a total of 1,972 cells (Supplementary Fig. 10a). Of these, 20 cells were found to be significantly contaminated with oligodendroglial cell transcripts (Supplementary Fig. 10b), leaving 1,952 cells for all downstream analyses. Quantification of ERCC spike-in control RNA indicated high accuracy and sensitivity of our single-cell profiling (Supplementary Fig. 10c). Genes were considered expressed if their expression was detected in more than 10 cells.

Our single-cell sequencing was not comprehensive, and with improved sequencing approaches further genetic subdivisions may be identified. Single-cell and pooled-cell dissections were not precisely matched, for example motor-projecting midline nuclei were not dissected for single-cell RNAseq. However, pooled-cell and single-cell RNAseq are in close agreement (Supplementary Fig. 10d), indicating that our results are robust to collection method.

Single-cell clustering and marker genes: Single-cell clusters were defined using the *Seurat* R package (version 2.1)[66,67]. Data were log transformed and scaled. For identifying variable genes, genes were divided into 20 bins based on average expression, and genes that were more than 1 standard deviation away from average dispersion within a bin were used for downstream analysis. Single-cell clustering was performed for all cells combined together (Fig. 5a–c) and separately for each projection system (Fig. 6) using shared nearest neighbor clustering and limiting the analysis to the top 10 principal components for distance calculation. Clusters were defined by the Louvain algorithm, clustering resolution was set to 0.6. Clusters of cells were visualized using t-distributed stochastic neighbor embedding (tSNE) using the top 10 principal components as input and perplexity set to 30. Marker genes for each cluster were required to be expressed in at least 80% of the cells in the

cluster, to have a p-value $<10^{-5}$ (likelihood ratio test), a log2 fold change > 0.5. Projection of single-cell data onto pooled-cell principal components was obtained by multiplying (dot product) log-transformed and scaled single-cell data by the pooled-cell principal component loadings. To quantify the cluster membership probability for each cell, we trained a random forest classifier using 5-fold cross-validation via the *caret* and *randomForest* R packages on the normalized expression value of the same set of variable genes that were used for clustering in *Seurat* and number of trees and number of random variables both set to 500. The cluster probability corresponds to the normalized vote for that cluster in the random forest classifier and is shown as an alpha value on the tSNE plot in Supplementary Fig. 12.

Statistics and Reproducibility—Statistical methods for pooled-cell and single-cell RNAseq analyses are described in the "Analysis Methods" section, and those for electrophysiological experiments in the "Electrophysiology" section. No statistical methods were used to pre-determine sample sizes, but our sample sizes were similar to or larger than those reported in previous publications[19,56,68–71]. Unless otherwise specified, all correlations are shown as scatter plots with a curve fitted by linear regression and summary statistics included the fraction of variance explained (R^2) and a p-value of the fit determined by two-sided Student's t-test. The corresponding t-values and degrees of freedom can be found in Source Data 2. For t-tests and ANOVA with post-hoc Tukey HSD tests, data distribution was assumed to be normal, but this was not formally tested. For the human microarray data analysis (Fig. 2e), the distribution of projected PC1 scores was borderline normal (Shapiro-Wilks test for normality: p=0.0635), and significance was determined using non-parametric tests. When multiple hypotheses tests were carried out simultaneously, the p-values were adjusted for multiple testing using the Benjamini-Hochberg method. In Fig. 3d and Supplementary Fig. 4 and 8, the p-values were indicated as a range (*: <0.05, **: <0.01, ***: <0.001, ****:<0.0001) for better figure readability. Exact p-values can be found in Source Data 2.

For all confocal images, the experiments were conducted at least twice with similar results, and representative images are shown. For Fig. 1b and 3c, the confocal images show representative examples of fluorescent labeling of cells that were otherwise used for data collection.

Reporting Summary—Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

References

1. Sherman SM & Guillery RW The role of the thalamus in the flow of information to the cortex. Philos. Trans. R. Soc. B Biol. Sci 357, 1695–1708 (2002).

2. Jones EG The thalamus. (Cambridge University Press, 2007).

3. Nakajima M & Halassa MM Thalamic control of functional cortical connectivity. Curr. Opin. Neurobiol 44, 127–131 (2017). [PubMed: 28486176]

4. Clascá F, Rubio-Garrido P & Jabaudon D Unveiling the diversity of thalamocortical neuron subtypes. Eur. J. Neurosci 35, 1524–32 (2012). [PubMed: 22606998]

5. Willis T Cerebri anatome, cui accessit nervorum descriptio et usus. (1664).

6. Jasper H Diffuse projection systems: the integrative action of the thalamic reticular system. Electroencephalogr. Clin. Neurophysiol 1, 405–19; discussion 419–20 (1949). [PubMed: 18421831]

7. Smith Y et al. The thalamostriatal system in normal and diseased states. Front. Syst. Neurosci 8, 5 (2014). [PubMed: 24523677]

8. Jones EG & Hendry SHC Differential Calcium Binding Protein Immunoreactivity Distinguishes Classes of Relay Neurons in Monkey Thalamic Nuclei. Eur. J. Neurosci 1, 222–246 (1989). [PubMed: 12106154]

9. Jones EG Viewpoint: the core and matrix of thalamic organization. Neuroscience 85, 331–45 (1998). [PubMed: 9622234]

10. Jones EG The thalamic matrix and thalamocortical synchrony. Trends Neurosci. 24, 595–601 (2001). [PubMed: 11576674]

11. Cowan WM & Powell TP A study of thalamo-striate relations in the monkey. Brain 79, 364–90 (1956). [PubMed: 13364089]

12. Kato S et al. Action Selection and Flexible Switching Controlled by the Intralaminar Thalamic Neurons. Cell Rep. 22, 2370–2382 (2018). [PubMed: 29490273]

13. Nelson SB, Sugino K & Hempel CM The problem of neuronal cell types: a physiological genomics approach. Trends Neurosci. 29, 339–345 (2006). [PubMed: 16714064]

14. Kepecs A & Fishell G Interneuron cell types are fit to function. Nature 505, 318–26 (2014). [PubMed: 24429630]

15. Murray KD, Choudary PV & Jones EG Nucleus- and cell-specific gene expression in monkey thalamus. Proc. Natl. Acad. Sci 104, 1989–1994 (2007). [PubMed: 17261798]

16. Frangeul L et al. A cross-modal genetic framework for the development and plasticity of sensory pathways. Nature 538, 96–98 (2016). [PubMed: 27669022]

17. Nagalski A et al. Molecular anatomy of the thalamic complex and the underlying transcription factors. Brain Struct. Funct 221, 2493–2510 (2016). [PubMed: 25963709]

18. Franklin KBJ & Paxinos G Paxinos and Franklin's The mouse brain in stereotaxic coordinates.

19. Sugino K et al. Mapping the transcriptional diversity of genetically and anatomically defined cell populations in the mouse brain. Elife 8, (2019).

20. Hawrylycz MJ et al. An anatomically comprehensive atlas of the adult human brain transcriptome. Nature 489, 391–399 (2012). [PubMed: 22996553]

21. Freeman SA, Desmazières A, Fricker D, Lubetzki C & Sol-Foulon N Mechanisms of sodium channel clustering and its influence on axonal impulse conduction. Cell. Mol. Life Sci 73, 723–35 (2016). [PubMed: 26514731]

22. Heinemann SH, Rettig J, Wunder F & Pongs O Molecular and functional characterization of a rat brain K $_v$ β 3 potassium channel subunit. FEBS Lett. 377, 383–389 (1995). [PubMed: 8549760]

23. Rudy B & McBain CJ Kv3 channels: voltage-gated K+ channels designed for high-frequency repetitive firing. Trends Neurosci. 24, 517–26 (2001). [PubMed: 11506885]

24. Okaty BW, Miller MN, Sugino K, Hempel CM & Nelson SB Transcriptional and electrophysiological maturation of neocortical fast-spiking GABAergic interneurons. J. Neurosci 29, 7040–52 (2009). [PubMed: 19474331]

25. Nakamura KC, Sharott A & Magill PJ Temporal coupling with cortex distinguishes spontaneous neuronal activities in identified basal ganglia-recipient and cerebellar-recipient zones of the motor thalamus. Cereb. Cortex 24, 81–97 (2014). [PubMed: 23042738]

26. Puil E, Meiri H & Yarom Y Resonant behavior and frequency preferences of thalamic neurons. J. Neurophysiol 71, 575–82 (1994). [PubMed: 8176426]

27. Fogerson PM & Huguenard JR Tapping the Brakes: Cellular and Synaptic Mechanisms that Regulate Thalamic Oscillations. Neuron 92, 687–704 (2016). [PubMed: 27883901]

28. Economo MN et al. A platform for brain-wide imaging and reconstruction of individual neurons. Elife 5, e10566 (2016). [PubMed: 26796534]

29. Bickford ME, Zhou N, Krahe TE, Govindaiah G & Guido W Retinal and Tectal 'Driver-Like' Inputs Converge in the Shell of the Mouse Dorsal Lateral Geniculate Nucleus. J. Neurosci. 35, 10523–34 (2015). [PubMed: 26203147]

30. Kuramoto E et al. Two types of thalamocortical projections from the motor thalamic nuclei of the rat: a single neuron-tracing study using viral vectors. Cereb. Cortex 19, 2065–77 (2009). [PubMed: 19174446]

31. Groh A et al. Convergence of Cortical and Sensory Driver Inputs on Single Thalamocortical Cells. Cereb. Cortex 24, 3167–3179 (2014). [PubMed: 23825316]

32. Lu E, Llano DA & Sherman SM Different distributions of calbindin and calretinin immunostaining across the medial and dorsal divisions of the mouse medial geniculate body. Hear. Res 257, 16–23 (2009). [PubMed: 19643174]

33. Mo C, Petrof I, Viaene AN & Sherman SM Synaptic properties of the lemniscal and paralemniscal pathways to the mouse somatosensory thalamus. Proc. Natl. Acad. Sci 114, E6212–E6221 (2017). [PubMed: 28696281]

34. Ramcharan EJ, Gnadt JW & Sherman SM Higher-order thalamic relays burst more than first-order relays. Proc. Natl. Acad. Sci 102, 12236–12241 (2005). [PubMed: 16099832]

35. Ferster D, Chung S & Wheat H Orientation selectivity of thalamic input to simple cells of cat visual cortex. Nature 380, 249–252 (1996). [PubMed: 8637573]

36. Van der Werf YD, Witter MP & Groenewegen HJ The intralaminar and midline nuclei of the thalamus. Anatomical and functional evidence for participation in processes of arousal and awareness. Brain Res. Brain Res. Rev 39, 107–40 (2002). [PubMed: 12423763]

37. Saalmann YB Intralaminar and medial thalamic influence on cortical synchrony, information transmission and cognition. Front. Syst. Neurosci 8, 83 (2014). [PubMed: 24847225]

38. Rikhye RV, Wimmer RD & Halassa MM Toward an Integrative Theory of Thalamic Function. Annu. Rev. Neurosci 41, 163–183 (2018). [PubMed: 29618284]

39. Guo K, Yamawaki N, Svoboda K & Shepherd GMG Anterolateral Motor Cortex Connects with a Medial Subdivision of Ventromedial Thalamus through Cell Type-Specific Circuits, Forming an Excitatory Thalamo-Cortico-Thalamic Loop via Layer 1 Apical Tuft Dendrites of Layer 5B Pyramidal Tract Type Neurons. J. Neurosci 38, 8787–8797 (2018). [PubMed: 30143573]

40. Bennett C et al. Higher-Order Thalamic Circuits Channel Parallel Streams of Visual Information in Mice. Neuron 102, 477–492.e5 (2019). [PubMed: 30850257]

41. Mandelbaum G et al. Distinct Cortical-Thalamic-Striatal Circuits through the Parafascicular Nucleus. Neuron (2019). doi:10.1016/j.neuron.2019.02.035

42. Suzuki-Hirano A et al. Dynamic spatiotemporal gene expression in embryonic mouse thalamus. J. Comp. Neurol 519, 528–543 (2011). [PubMed: 21192082]

43. Wong SZH et al. In vivo clonal analysis reveals spatiotemporal regulation of thalamic nucleogenesis. PLoS Biol. 16, e2005211 (2018). [PubMed: 29684005]

44. Altman J & Bayer SA Development of the diencephalon in the rat. V. Thymidine-radiographic observations on internuclear and intranuclear gradients in the thalamus. J. Comp. Neurol 188, 473–99 (1979). [PubMed: 489804]

45. Angevine JB Time of neuron origin in the diencephalon of the mouse. An autoradiographic study. J. Comp. Neurol 139, 129–187 (1970). [PubMed: 5463599]

46. Shi W et al. Ontogenetic establishment of order-specific nuclear organization in the mammalian thalamus. Nat. Neurosci 20, 516–528 (2017). [PubMed: 28250409]

47. Grant E, Hoerder-Suabedissen A & Molnár Z The Regulation of Corticofugal Fiber Targeting by Retinal Inputs. Cereb. Cortex 26, 1336–1348 (2016). [PubMed: 26744542]

48. Antón-Bolaños N, Espinosa A & López-Bendito G Developmental interactions between thalamus and cortex: a true love reciprocal story. Curr. Opin. Neurobiol 52, 33–41 (2018). [PubMed: 29704748]

49. Cembrowski MS & Menon V Continuous Variation within Cell Types of the Nervous System. Trends Neurosci. 41, 337–348 (2018). [PubMed: 29576429]

50. Kuramoto E et al. Complementary distribution of glutamatergic cerebellar and GABAergic basal ganglia afferents to the rat motor thalamic nuclei. Eur. J. Neurosci 33, 95–109 (2011). [PubMed: 21073550]

References (Methods-only)

51. Tervo DGR et al. A Designer AAV Variant Permits Efficient Retrograde Access to Projection Neurons. Neuron 92, 372–382 (2016). [PubMed: 27720486]

52. Kita T, Shigematsu N & Kita H Intralaminar and tectal projections to the subthalamus in the rat. Eur. J. Neurosci 44, 2899–2908 (2016). [PubMed: 27717088]

53. Mandelbaum G et al. Distinct Cortical-Thalamic-Striatal Circuits Through the Parafascicular Nucleus. bioRxiv 370734 (2018). doi:10.1101/370734

54. Hempel CM, Sugino K & Nelson SB A manual method for the purification of fluorescently labeled neurons from the mammalian brain. Nat. Protoc 2, 2924–2929 (2007). [PubMed: 18007629]

55. Cembrowski MS et al. Spatial Gene-Expression Gradients Underlie Prominent Heterogeneity of CA1 Pyramidal Neurons. Neuron 89, 351–68 (2016). [PubMed: 26777276]

56. Cembrowski MS et al. Dissociable Structural and Functional Hippocampal Outputs via Distinct Subiculum Cell Classes. Cell 173, 1280–1292.e18 (2018). [PubMed: 29681453]

57. Picelli S et al. Smart-seq2 for sensitive full-length transcriptome profiling in single cells. Nat. Methods 10, 1096–1098 (2013). [PubMed: 24056875]

58. Soumillon M, Cacchiarelli D, Semrau S, Oudenaarden A van & Mikkelsen TS Characterization of directed differentiation by high-throughput single-cell RNA-Seq. bioRxiv 003236 (2014). doi: 10.1101/003236

59. Murphy SD et al. The janelia workstation for neuroscience. Keystone Big Data Biol. San Fr. CA (2014).

60. Johnson H, Harris G, Williams K & others. BRAINSFit: mutual information rigid registrations of whole-brain 3D images, using the insight toolkit. Insight J 57, (2007).

61. Ng L et al. An anatomic gene expression atlas of the adult mouse brain. Nat. Neurosci 12, 356–362 (2009). [PubMed: 19219037]

62. Dobin A et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29, 15–21 (2013). [PubMed: 23104886]

63. Robinson MD, McCarthy DJ & Smyth GK edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26, 139–140 (2010). [PubMed: 19910308]

64. Anders S & Huber W Differential expression analysis for sequence count data. Genome Biol. 11, R106 (2010). [PubMed: 20979621]

65. Love MI, Huber W & Anders S Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 15, (2014).

66. Satija R, Farrell JA, Gennert D, Schier AF & Regev A Spatial reconstruction of single-cell gene expression data. Nat. Biotechnol 33, 495–502 (2015). [PubMed: 25867923]

67. Butler A, Hoffman P, Smibert P, Papalexi E & Satija R Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nat. Biotechnol 36, 411–420 (2018). [PubMed: 29608179]

68. Cembrowski MS, Wang L, Sugino K, Shields BC & Spruston N Hipposeq: a comprehensive RNA-seq database of gene expression in hippocampal principal neurons. Elife 5, e14997 (2016). [PubMed: 27113915]

69. Miller MN, Okaty BW & Nelson SB Region-specific spike-frequency acceleration in layer 5 pyramidal neurons mediated by Kv1 subunits. J. Neurosci 28, 13716–26 (2008). [PubMed: 19091962]

70. Miller MN, Okaty BW, Kato S & Nelson SB Activity-dependent changes in the firing properties of neocortical fast-spiking interneurons in the absence of large changes in gene expression. Dev. Neurobiol 71, 62–70 (2011). [PubMed: 21154910]

71. Dani VS et al. Reduced cortical activity due to a shift in the balance between excitation and inhibition in a mouse model of Rett syndrome. Proc. Natl. Acad. Sci. U. S. A 102, 12560–5 (2005). [PubMed: 16116096]
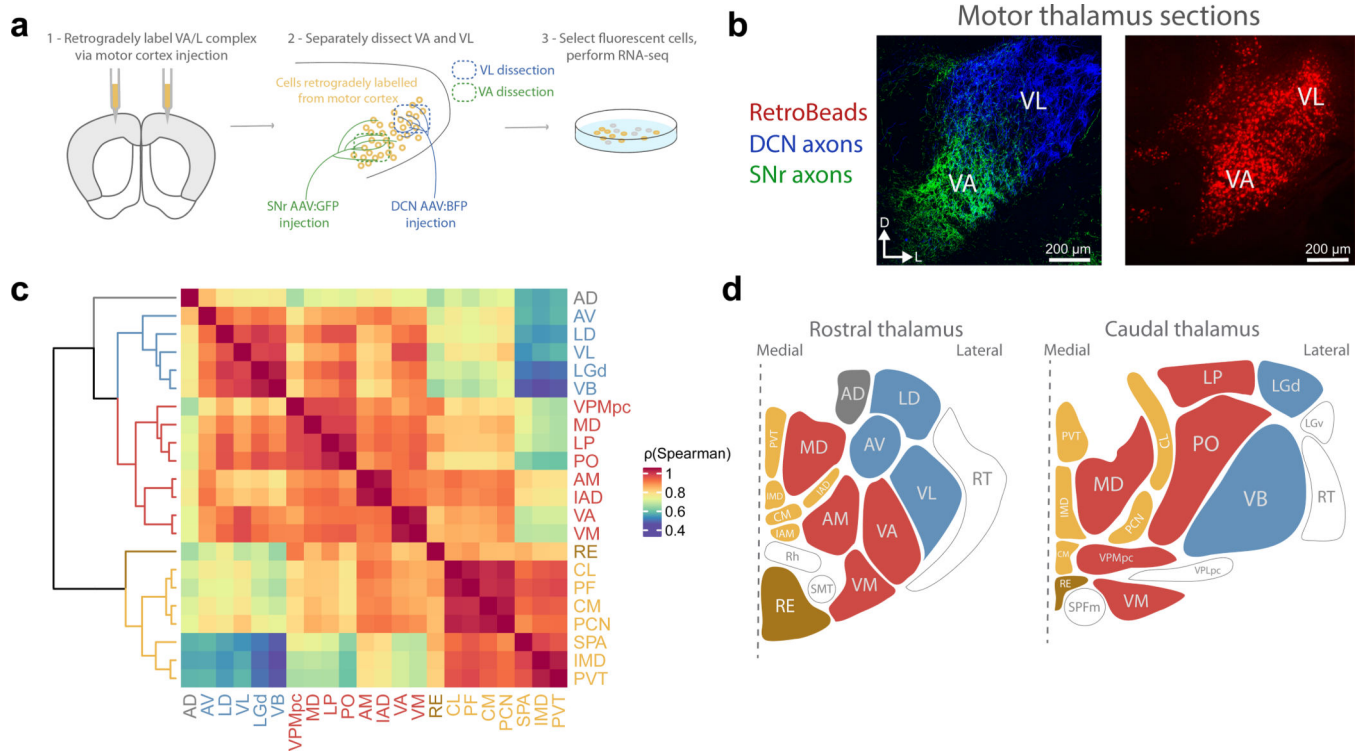
Fig. 1 –. Projection-specific RNAseq approach and unbiased clustering of thalamic gene expression profiles.

a. Schematic of experimental pipeline to obtain transcriptomic atlas of the thalamus. In this example, motor thalamic neurons were retrogradely labeled from their primary projection target (motor cortex), manually dissected, and sorted. Viruses expressing green and blue fluorescent proteins (GFP, BFP), respectively, were injected to the deep cerebellar nuclei (DCN) and substantia nigra pars reticulata (SNr) to label motor nuclear subdivisions, ventral lateral (VL) and ventral anterior (VA), respectively[25,50].

b. Example labelling from scheme shown in Fig. 1a. Coronal sections. Scale bars = 200 μm.

c. Hierarchical clustering of thalamic nuclei using Spearman's correlation of the top 500 most differentially expressed genes (Supplementary Fig. 1c) across all 22 nuclei (total n=120 samples; using mean of 3–8 replicates per nucleus; see Source Data 1 for detailed information on each sample). Major profiles defined as the top 5 branches of cluster dendrogram. Acronyms of nuclei are defined in Supplementary Table 1.

d. Topographic localization of gene expression profiles in thalamus. Coronal thalamic section schematics with nuclei colored as in Fig. 1c. Unsampled nuclei are left uncolored.
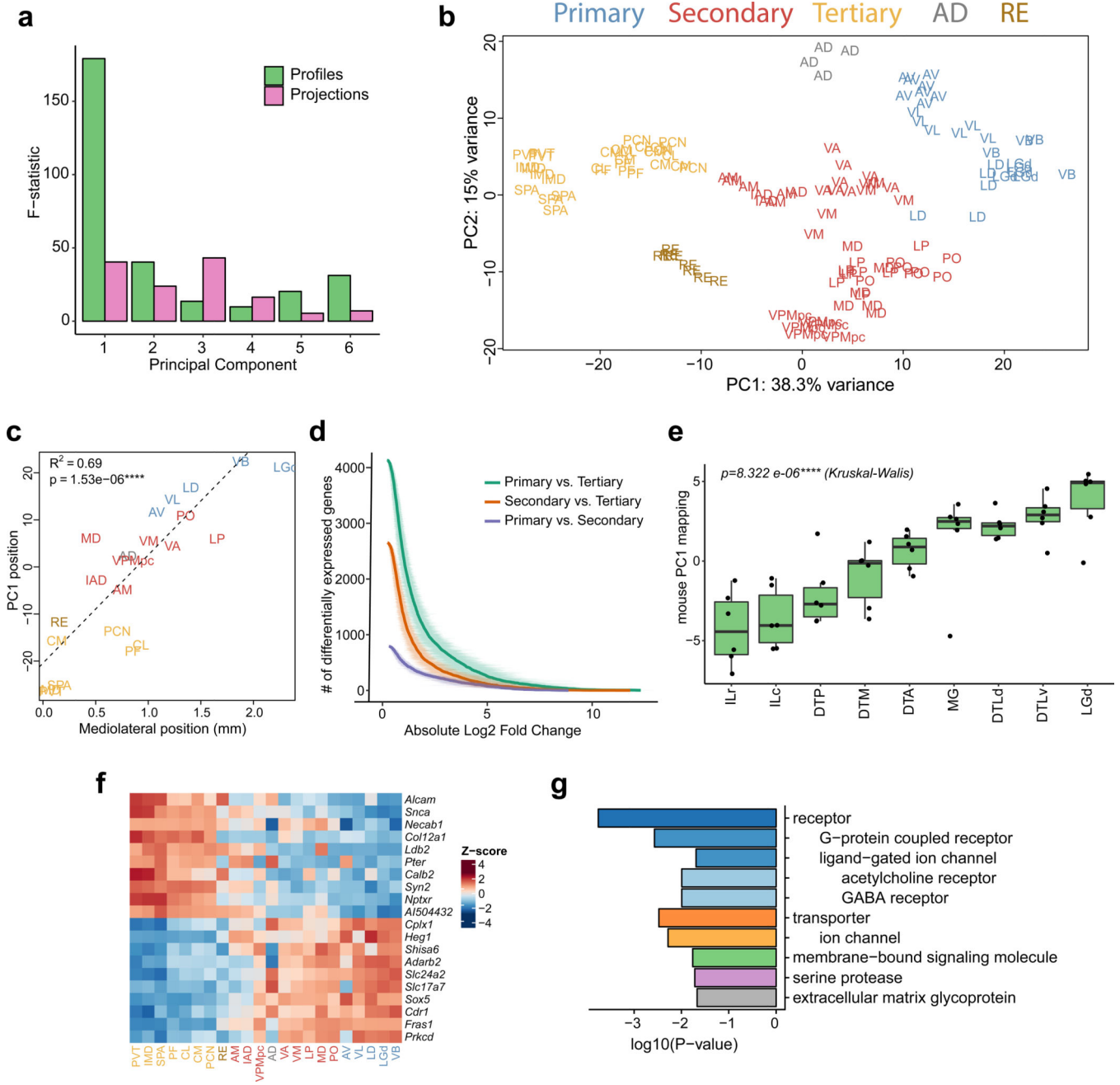
Fig. 2 –. Principal component analysis reveals gene expression differences conserved across thalamic projection systems.

a. Bar plot showing how the principle components of gene expression relate to the profiles from hierarchical clustering ('Profiles'; Fig. 1c) and projection targets ('Projections'), assessed by F-statistic (ratio of between- and within-group variance). This analysis included all samples retrogradely labeled from the neocortex (n=90 samples).

b. PCA showing separation of functional nuclear profiles in the first two principal components. The underlying gene set, sample set, and color scheme are the same as in Fig. 1c.

c. PC1 position is highly correlated (as determined by linear regression) with mediolateral position of the nuclei. Mediolateral positions are based on z voxel coordinates of nuclei centers in the Common Coordinate Framework of the Allen Mouse Brain Atlas. Included samples and nuclei coloring scheme is the same as in Fig. 2b.

d. Differential gene expression is larger between primary and tertiary nuclei than between primary and secondary or secondary and tertiary nuclei. Plot shows the number of differentially expressed genes at each log fold change for the three comparisons level (shown as mean ± standard deviation of bootstrapped value; 28 primary, 50 secondary, and 30 tertiary samples were subsampled with replacement to the size of the smallest group).

e. Projection of human microarray data onto mouse PC1 using the mouse gene loadings of matching homologous genes. LGd (primary profile) mapped onto the positive side and intralaminar nuclei (tertiary profile) on the negative end of the axis (Kruskal-Wallis test for differences between all 9 thalamic areas: $p = 8.3 \times 10^{-6}$). Boxes show median and quartiles, and whiskers extend to up to the highest and lowest value that is within the 1.5x interquartile range from the upper and lower quartiles. Individual samples (n=6 for each region) are shown as black dots. Acronyms for the human thalamic regions are listed in Supplementary Table 4.

f. Heatmap of genes with strongest positive and negative loadings on PC1. Mean expression per nucleus (22 nuclei, n=120 samples) is shown. Nuclei are ordered by their mean position on PC1 of Fig. 2a. Colors represent gene-wise Z-scores.

g. Genes relevant to neurotransmission are overrepresented among genes with high PC1 loadings in our dataset. The ten most highly overrepresented PANTHER protein class terms are shown. p-values based on two-sided Fisher's exact test for each term. Overrepresentation in the top 200 genes with the highest absolute PC1 loading was assessed. The background gene pool included all expressed genes (n=16,538). Indentation indicates gene subfamily.
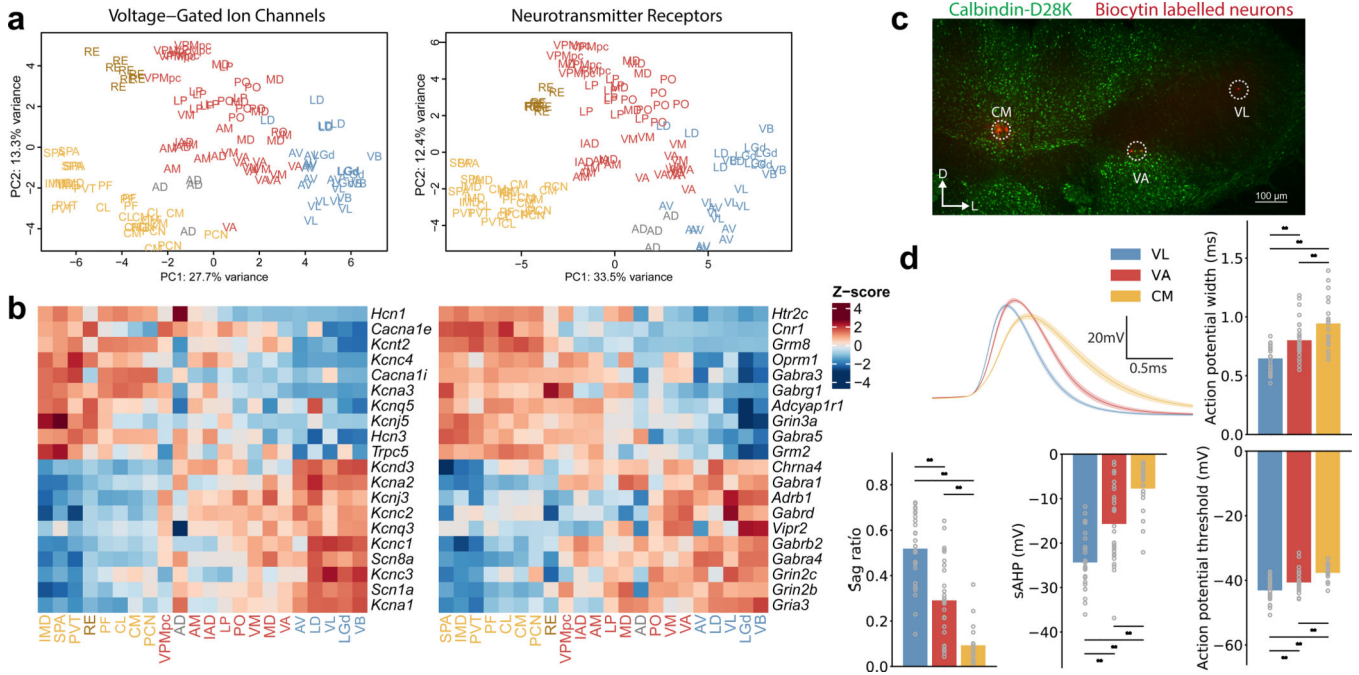
Fig. 3 –. Systematic variation of functionally relevant genes and electrophysiological properties across thalamic nuclear profiles.

a. PCA including only genes encoding voltage-gated ion channels and neurotransmitter/ neuromodulator receptors. Sample set and colors are as in Fig. 1c.

b. Heatmap for genes with the highest gene loadings in PC1 from Fig. 3a. Voltage-gated ion channels on the left and neurotransmitter receptors on the right. Mean expression per nucleus is shown (22 nuclei, n=120 samples). Colors represent gene-wise Z-scores.

c. Representative example of neurons labeled during whole-cell patch-clamp recordings from VL (primary profile), VA (secondary profile), and CM (tertiary profile). Neurons were labeled with biocytin (red) and localized to individual nuclei with the aid of Calbindin-D28K (*Calb1*) immunolabeling (green). Scale bar = 100 μm.

d. Systematic variation of electrophysiological properties across profiles. Upper left shows average action potential shape for VL, VA, and CM neurons (mean ± standard error of the mean). Remaining panels show comparisons for four physiological measurements across these nuclei (one-way ANOVA with post-hoc Tukey HSD test, all comparisons p<0.01; for exact p-values see Source Data 2). Sample contained 29 VL neurons, 30 VA neurons, and 20 CM neurons.
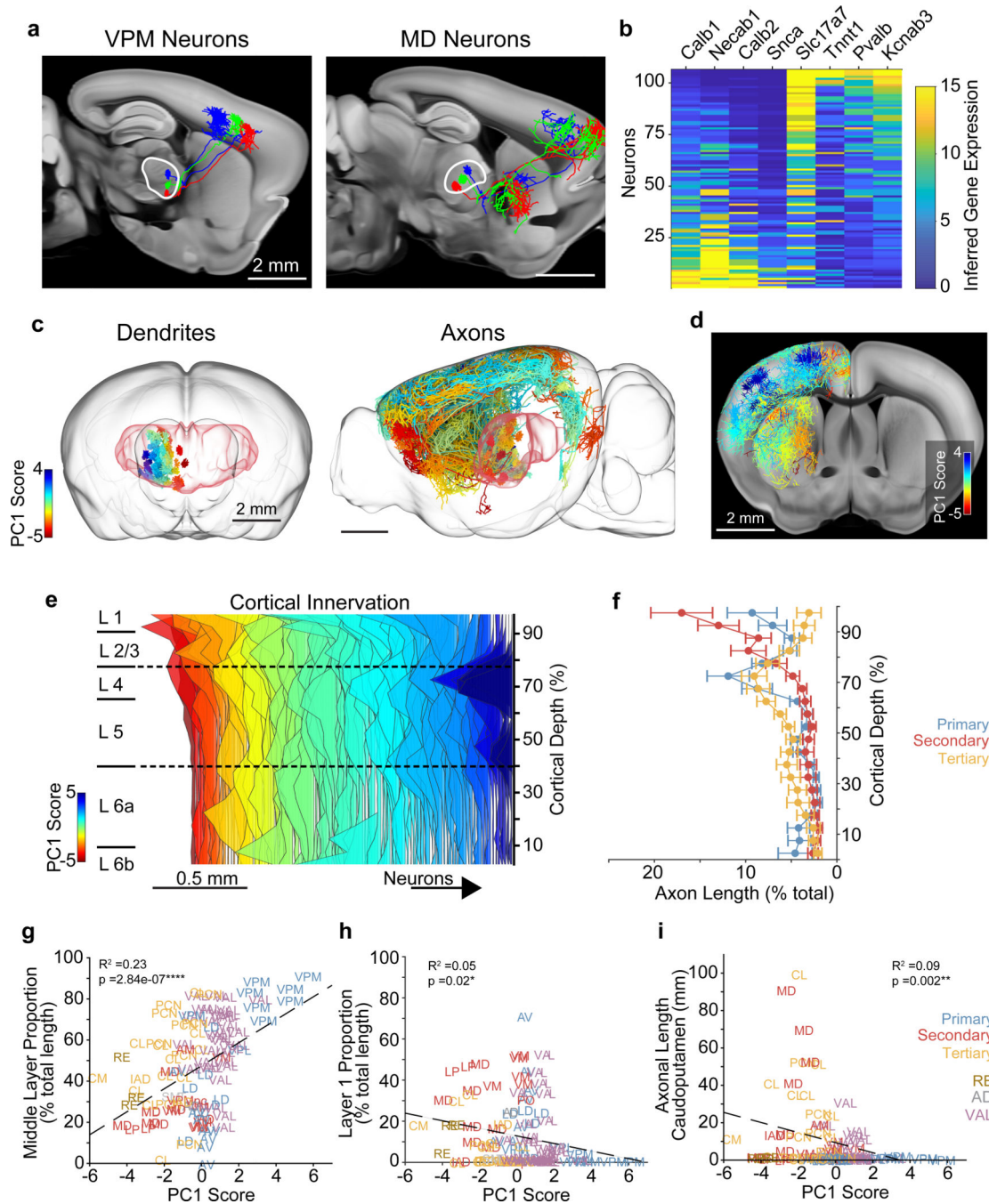
Fig. 4 –. Single thalamic neuron reconstructions show covariation of axonal morphology and gene expression.

a. Example of three representative neurons in the ventral posteromedial nucleus (VPM) and the mediodorsal nucleus (MD) shown in a sagittal view. Overall 106 thalamic neurons were reconstructed.

b. Heatmap of inferred gene expression for all reconstructed neurons. Gene expression was calculated for each neuron by taking the gene expression energy score from the Allen Institute's In Situ Hybridization Atlas at the soma location (see methods).

c. 3D visualization of the dendritic (left) and axonal morphology (right) of all reconstructed thalamic neurons (n=106). Colors represent the gene expression PC1 score for each neuron. The position of the thalamus is highlighted in red.

d. Coronal cross-section of same data shown in right panel of Fig. 4c.

e. Layer preference of thalamic axons in the cortex varies along PC1. Each area plot shows the binned axonal length along the entire cortical depth of one neuron. 100% is the pial surface and 0% is the white matter surface. Shown on the left is the approximate location of the different cortical layers. Each neuron's density is colored by its position on PC1. The dashed line represents the middle layer used in Fig. 4g. Scale bar is 0.5 mm (axonal length in bin).

f. Axonal length by cortical depth for primary (n=24 neurons), secondary (n=21 neurons), and tertiary (n=21 neurons) thalamic nuclei, normalized by total axonal length for each neuron, binned into 20 bins. Dots and error bars represent means ± standard errors of the mean. Cortical depth scale as in Fig. 4e.

g. Relationship between the gene expression PC1 score and the proportion of a neuron's cortical axonal length in the middle layers of the cortex. Primary thalamic neurons have a larger proportion of their cortical axons in the middle layers of the cortex.

h. Relationship between the gene expression PC1 score and the proportion of a neuron's cortical axonal length in layer 1 of the cortex. Secondary thalamic neurons have a larger proportion of their cortical axons in layer 1 of the cortex

i. Relationship between a neuron's gene expression PC1 score and its total axonal length within the caudoputamen. Subcompartments of MD in single-cell reconstructions are likely unresolvable. Color key applies to Fig. 4g-i. All correlations are determined via linear regression and p-values are calculated via two-sided Student's t-test. All intensity images and 3D anatomical models are taken from the Allen Mouse Brain Atlas. Scale bars are 2 mm unless stated otherwise. Reported distances and positions are based on the Allen Institute's Common Coordinate Framework.
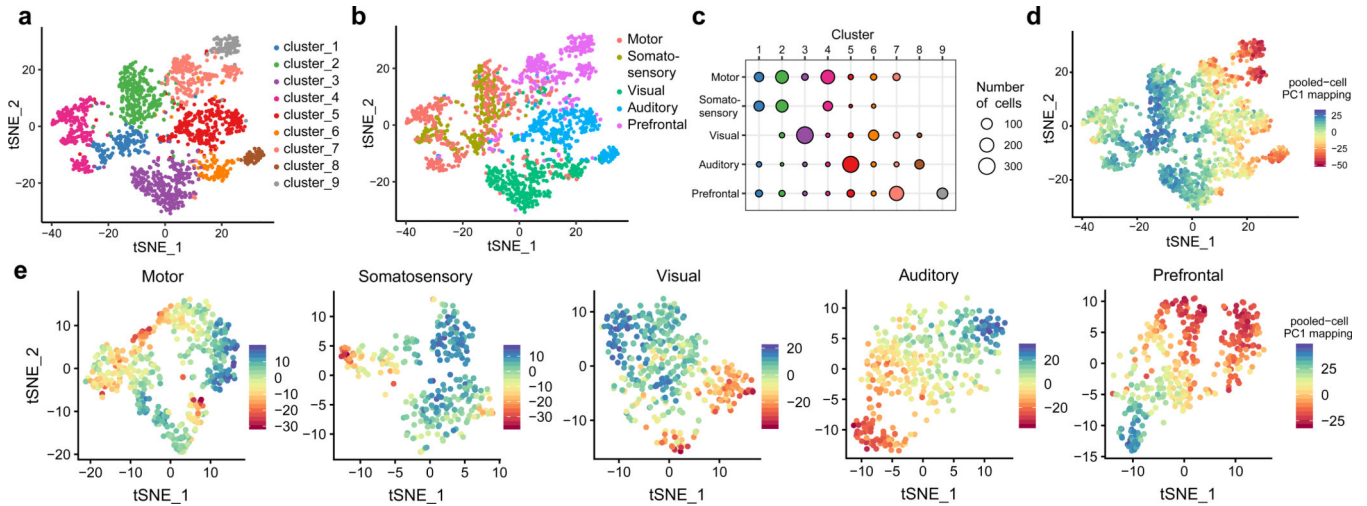
Fig. 5 –. Combined influence of projection and profile on single cell identity

a. tSNE plot showing output of unsupervised single-cell clustering after combining neurons of all five projection types.

b. tSNE plot as in Fig. 5a colored by cortical projection target area.

c. Bubble plot showing distribution of clusters across cortical projection target types. All clusters contain cells from several projections, with the exception of cluster 9.

d. tSNE plot of all cells, colored by their position on pooled-cell PC1 (from Fig. 2a).

e. tSNE plot for each individual projection system, colored by their position on pooled-cell PC1 (from Fig. 2a). For each projection target, only the major projection complex in thalamus was collected, avoiding the relatively sparsely labeled midline cells, which would impede cell sorting if collected. This likely explains lack of full PC1 coverage in all projection systems. The number of cells analyzed per projection area were 477 for motor, 282 for somatosensory, 429 for visual, 376 for auditory, and 388 for prefrontal.
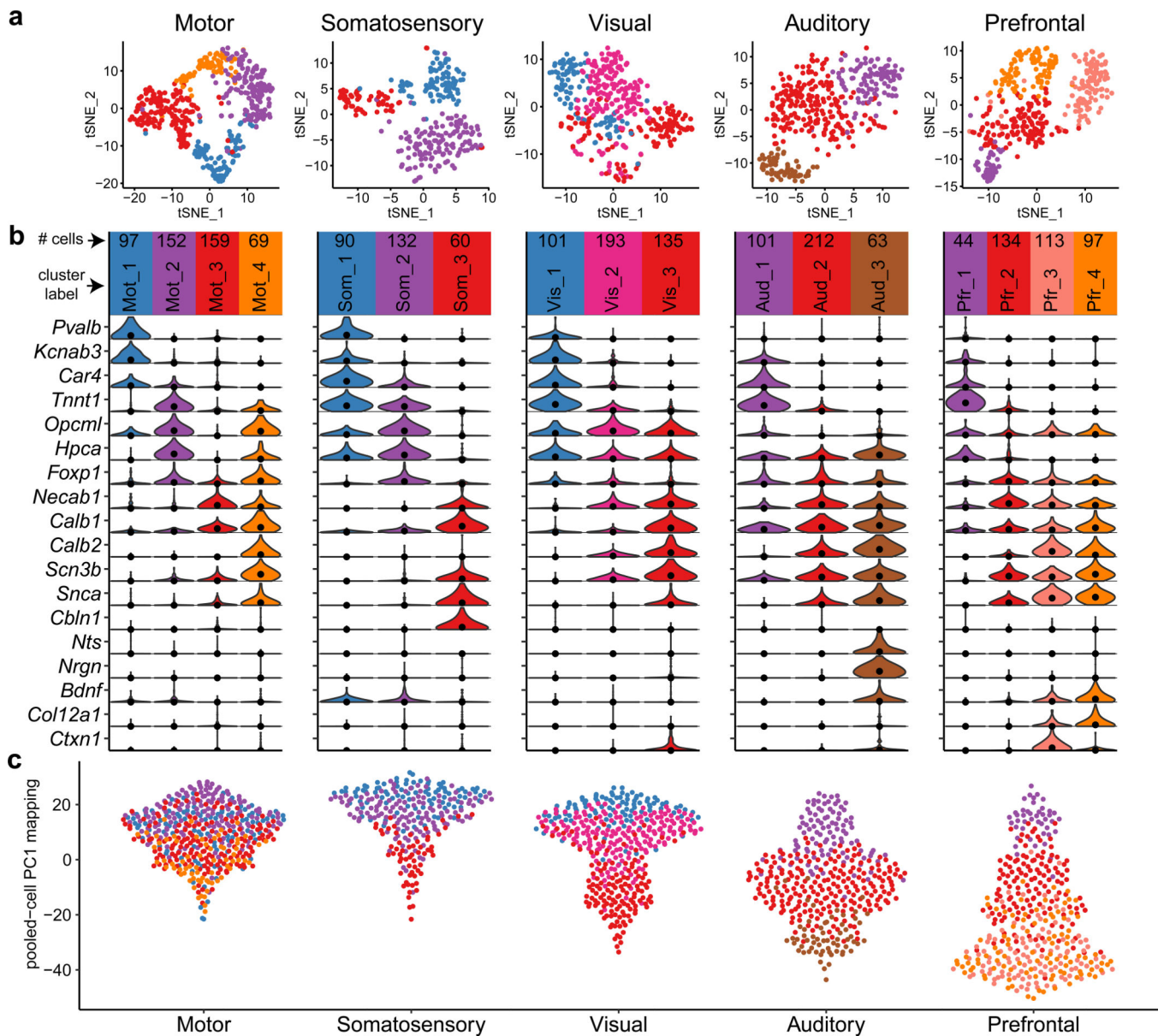
Fig. 6 –. Genetic basis of the single-cell clusters of the major thalamic projection systems.
a. Overview of single-cell clusters within each projection system visualized via tSNE. Cells colored by cluster identity. The cells included in this figure are the same as in Fig. 5.
b. Violin plots for cluster marker genes in each projection system (inclusion criteria are likelihood ratio test p-value $< 10^{-5}$ and log2 fold change > 0.5 for each cluster). Expression values are counts per million. The y-axis is rescaled for each gene and each projection type. Black dots indicate median.
c. Mapping of single-cell data onto pooled-cell PC1 from Fig. 2a. Each dot is a single cell colored by cluster identity as in Fig. 6a.
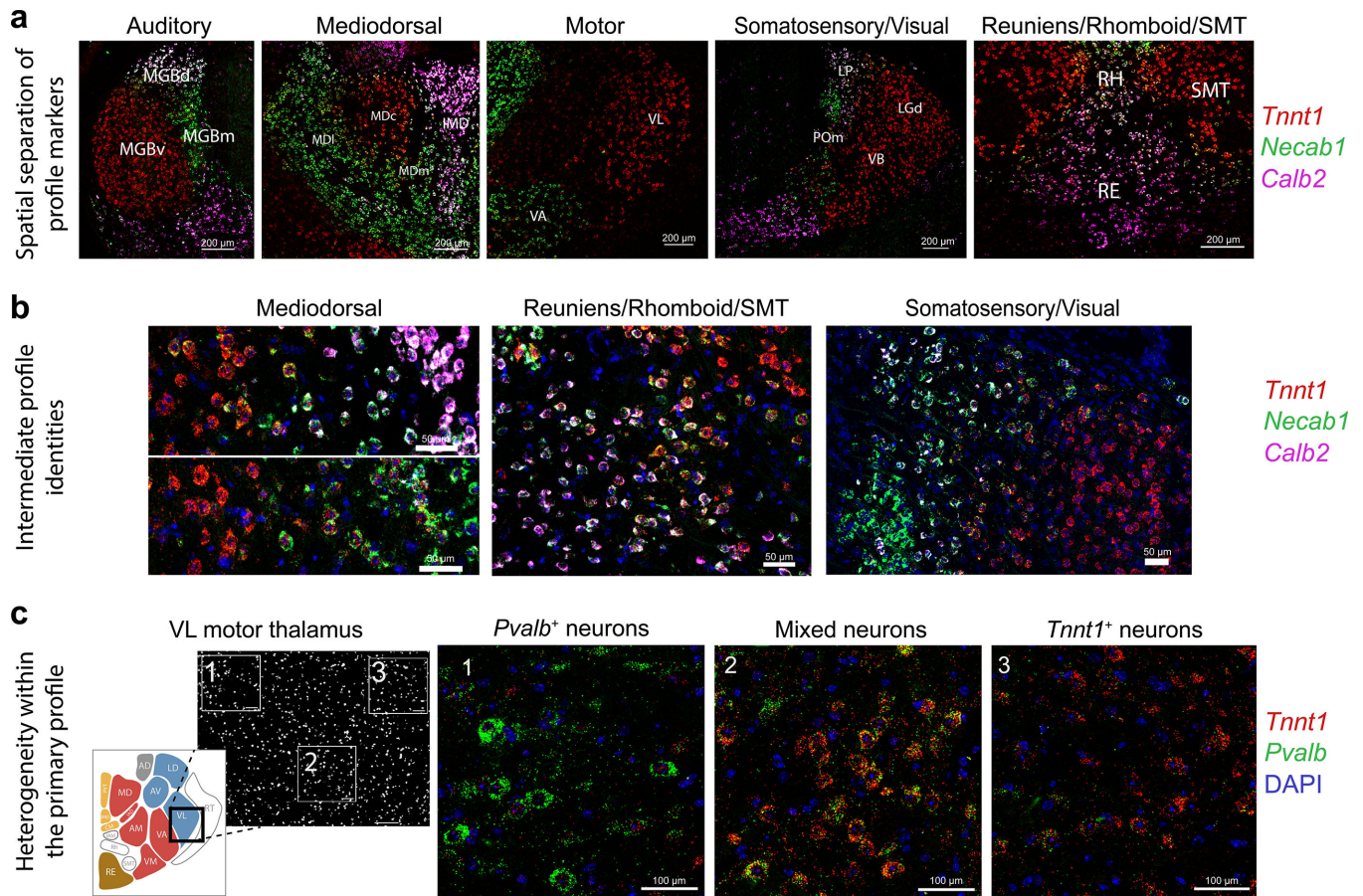
Fig. 7 –. Multi-FISH confirms continuous, topographic variation between profiles

a. Topographic distribution of marker genes within 6 major anatomical groups. Multi-FISH with probes for *Calb2* (pink), *Tnnt1* (red), and *Necab1* (green). See Supplementary Fig. 13 and 14 for expanded views and quantification. All multi-FISH experiments were repeated twice to ensure reproducibility. Coronal sections, with scale bars = 200 μm.

b. Intermediate cell types expressing two marker genes were present at nuclear boundaries. Coronal sections, with scale bars = 50 μm. Enlarged view and additional images in Supplementary Fig. 13.

c. Multi-FISH within VL thalamus comparing genes marking clusters in single-cell data. Left panel shows field of view (coronal section, scale bar = 100 μm). Right three panels (scale bars = 50 μm) show expansion of the three boxed areas, moving left to right. Middle box shows intermediate cells expressing both single-cell cluster markers. (red = *Tnnt1*, green = *Pvalb*, blue = DAPI).