A dual-strategy expression screen for candidate connectivity labels in the developing thalamus

Olivia Bibollet-Bahena[1], Tatsuya Okafuji[1], Karsten Hokamp[1] and Kevin J. Mitchell[1,2]

[1]Smurfit Institute of Genetics

[2]Institute of Neuroscience

Trinity College Dublin

Dublin 2, Ireland

Address for correspondence: Kevin.Mitchell@tcd.ie

Abstract

The thalamus or "inner chamber" of the brain is divided into ~30 discrete nuclei, with highly specific patterns of afferent and efferent connectivity. To identify genes that may direct these patterns of connectivity, we used two strategies. First, we used a bioinformatics pipeline to survey the predicted proteomes of nematode, fruitfly, mouse and human for extracellular proteins containing any of a list of motifs found in known guidance or connectivity molecules. Second, we performed clustering analyses on the Allen Developing Mouse Brain Atlas data to identify genes encoding surface proteins expressed with temporal profiles similar to known guidance or connectivity molecules. In both cases, we then screened the resultant genes for selective expression patterns in the developing thalamus. These approaches identified 82 candidate connectivity labels in the developing thalamus. These molecules include many members of the Ephrin, Eph-receptor, cadherin, protocadherin, semaphorin, plexin, Odz/teneurin, Neto, cerebellin, calsyntenin and Netrin-G families, as well as diverse members of the immunoglobulin (Ig) and leucine-rich receptor (LRR) superfamilies, receptor tyrosine kinases and phosphatases, a variety of growth factors and receptors, and a large number of miscellaneous membrane-associated or secreted proteins not previously implicated in axonal guidance or neuronal connectivity. The diversity of their expression patterns indicates that thalamic nuclei are highly differentiated from each other, with each one displaying a unique repertoire of these molecules, consistent with a combinatorial logic to the specification of thalamic connectivity.

Keywords: Sdk2, Kirrel3, Igsf9b, MDGA1, Syncam, CD47

Introduction

The thalamus, or "inner chamber" of the brain, is a crucial nexus in the brain's circuitry. It is not only a relay station that conveys sensory information from the periphery to the cerebral cortex, it is also a conduit for cortico-cortical communication [1], as well as a central node in pathways controlling action selection, through cortico-baso-thalamo-cortical loops [2]. In addition, the thalamus is interconnected with many other brain structures, including hippocampus, hypothalamus, amygdala, inferior and superior colliculus [3], cerebellar nuclei [4], substantia nigra, brainstem, spinal cord and many others. The thalamus proper, or dorsal thalamus, is also intimately interconnected with the prethalamic reticular nucleus [5], which provides inhibitory regulation of information flow through thalamus.

A striking characteristic of the thalamus is its subdivision into ~30 discrete nuclei, which subserve distinct functions and which have highly selective connectivity patterns with the structures mentioned above, most obviously with specific cortical areas. Some nuclei, such as those conveying primary sensory information, project quite selectively to one or a small number of cortical areas, while others, which integrate signals from multiple sources, project in a more diffuse manner across larger areas of cortex. Within each nucleus there are also varying proportions of distinct cell types that either project to specific areas, driving receptive field properties in input layers of cortex, such as layer 4, or that project more widely across cortex and provide modulatory inputs, for example to layer 1 [6,7].

Different thalamic nuclei thus act as discrete targets for innervation from a wide array of sources, and, in turn, project their axons to numerous other structures with areal, laminar and cell-type specificity. In order to establish these connections, growing axons must be guided to appropriate regions, must recognise the appropriate target and avoid inappropriate ones, and must make the right kinds of synapses on the right kinds of cells. These processes are mediated in general by surface and secreted proteins, which act as signals and receptors, enabling cellular recognition for pathfinding and target selection.

Numerous protein families have been identified as playing important roles in thalamic axon guidance and connectivity. Many of these, including Ephrins/Eph-receptors, Netrins/DCC/Unc5s, Slits/Robos, Neuregulin-1/Erbb4, secreted semaphorins/

neuropilins and L1 cell adhesion molecules, mediate general processes such as avoidance of the hypothalamus, projection into the internal capsule and topographic organisation of thalamocortical and/or corticothalamic projections through this intermediate target region [reviewed in 8,9]. By contrast, very few molecules have been found so far that mediate more specific connectivity relationships of particular thalamic nuclei, such as Cdh6 [10], or that control sub-organisation of projections within thalamic nuclei, such as Lrp8 [11] or Ten-m3/Odz3 [12], which also regulates topography of projections to striatum [13].

At earlier stages, a number of studies have described the combinatorial expression patterns of patterning molecules or transcription factors across the thalamus, many of which are expressed selectively in some nuclei and not others [14-18]. Some have also documented the differential expression of selected surface proteins across the thalamus [15,19,20], and in some cases direct links have been shown between transcription factor and surface molecule expression [21,22]. However, these analyses have not been performed in a systematic or comprehensive fashion. Thus, while we have learned a lot about how the developing thalamus is patterned and how the fates of different nuclei are specified, we still know relatively little about the combinatorial code of surface molecules that specifies nuclear connectivity.

Here, we describe two parallel approaches to identify genes encoding surface or secreted proteins expressed in discrete patterns across the developing thalamus. First, we screened the predicted proteomes of human, mouse, fly and worm for conserved genes encoding predicted transmembrane proteins with any of a number of protein motifs commonly found in axon guidance molecules. These genes were screened by *in situ* hybridization to find those expressed in selective or differential patterns across the neonatal thalamus. Second, we analysed the Allen Developing Mouse Brain Atlas (devABA) database for genes encoding surface or secreted proteins, which showed a similar temporal expression profiles similar to known genes for axon guidance or synaptic connectivity. The expression patterns of these genes were then examined to identify selectively or differentially expressed genes at mid or late embryonic stages. Together, these approaches have identified 82 genes encoding candidate connectivity labels in the developing thalamus. The expression patterns of these genes are highly diverse, such that individual thalamic nuclei express distinct repertoires of these

surface molecules, consistent with a combinatorial logic to the specification of connectivity.

Results

A bioinformatics and expression screen to identify conserved candidate connectivity labels

The proteomes of mammals contain many predicted proteins of unknown function. We were interested to discover genes encoding predicted transmembrane proteins that contain any of a number of protein motifs found in known axon guidance molecules. As many known axon guidance molecules are highly conserved from vertebrates to invertebrates, we further concentrated on proteins that had predicted orthologues in the fruitfly Drosophila melanogaster and/or in the nematode Caenorhabditis elegans.

To identify such proteins, we used protein localisation and motif searches to annotate the predicted proteomes of worm, fly, mouse and human. We also generated a matrix of pairwise BLAST scores across all the members of these proteomes and clustered them using the TRIBE-MCL algorithm, as previously described [23]. We screened through the resultant outputs to find clusters with mammalian and fly or worm members encoding predicted transmembrane proteins, which also contained any of the following motifs: immunoglobulin (Ig) domain, fibronectin type 3 domain, cadherin domain, leucine-rich repeat (LRR), EGF repeat, CUB domain, sema domain, or plexin repeat. We were particularly interested in discovering novel axon guidance or connectivity cues and so excluded genes or gene families where such a function had already been demonstrated in the mouse at the time the screen was performed. Using this approach, we prioritized a set of 42 genes for expression screening (Table S1).

We performed *in situ* hybridization for these genes on P0 mouse brain sections, as previously described for genes encoding LRR proteins [23]. Here we concentrate on members of other gene families that show selective or differential expression across the thalamus (Table 1 Column A).

A. Conserved candidate connectivity labels	B. devABA candidate connectivity labels		
Clstn1	Alcam	EfnA2	Mdga1
Clstn2	Astn2	EfnA5	Nrn
Igsf9b	Bmp3	EfnB3	Ntng1
Kirrel3	Cadm1	EphA1	Ntng2
Neto1	Cbln2	EphA3	Ntrk2
Neto2	Cbln4	EphA4	Ntrk3
Odz1	Cd47	EphA6	Pcdh1
Odz2	Cdh2	EphA8	Pcdh10
Odz3	Cdh4	EphA10	Pcdh11X
Odz4	Cdh6	EphB1	Pcdh19
Sdk2	Cdh7	EphB2	Pcdh21
	Cdh8	EphB6	PlxnA1
	Cdh9	Fat3	PlxnA2
	Cdh10	Flt3	PlxnC1
	Cdh11	Fzd7	Ptpru
	Cdh12	Gfra1	Ret
	Cdh13	Gfra2	Rtn4rl1
	Cdh24	Gpc3	Sema3F
	Clstn1	Igfbp5	Sema6A
	Clstn2	Inhba	Sema7A
	Cntn6	Kit	Odz3
	Cntnap4	Lgi2	Tgfb2
	Dlk1	Lrp8	Trp53i11
	Dner	Lrrn3	Vgf
		Lypd1	Wif1

Table 1. Candidate connectivity labels. This table lists all 82 genes for which we present expression patterns. Clstn1, Clstn2 and Odz3 appear on both lists.

Immunoglobulin superfamily members: We found three Ig superfamily members with selective expression across the thalamus: Igsf9b, Kirrel3, and Sdk2 (Figure 1; Figure S1). Igsf9b (Immunoglobulin Superfamily, Member 9B) is an orthologue of the Drosophila gene *turtle*, which has been implicated in various aspects of axon guidance [reviewed in 24]. Both Igsf9b and Igsf9 have recently been shown to play roles in inhibitory synapse development in mammals [25,26]. Igsf9b is more strongly expressed in some thalamic nuclei than others, but we detected only weak expression of Igsf9 in thalamus (not shown).

Kirrel3 (Kin Of IRRE Like 3 (Drosophila), also known as NEPH2) is an orthologue of the fly gene *irregular chiasm* (*IrreC*), which functions in axonal pathfinding and

target selection [reviewed in 27] and the C. elegans gene *SYG-1*, which is implicated in selective synaptogenesis [28]. Kirrel3 has recently been implicated in axon guidance in the vomeronasal system [29] and synapse formation in the hippocampus [30], but functions in the developing thalamus have not been described. Kirrel1 and Kirrel2 both showed only weak/background expression in thalamus (not shown).

Sdk2 (Sidekick2) is an orthologue of the fly gene *sidekick*, which regulates cellular differentiation in the fly eye [31]. An important role for Sidekicks has also been demonstrated in the vertebrate retina in specifying lamina formation and synaptic connectivity [32,33]. Sdk2 is quite selectively expressed across thalamic nuclei while Sdk1 is expressed at high levels across the whole dorsal thalamus (not shown).
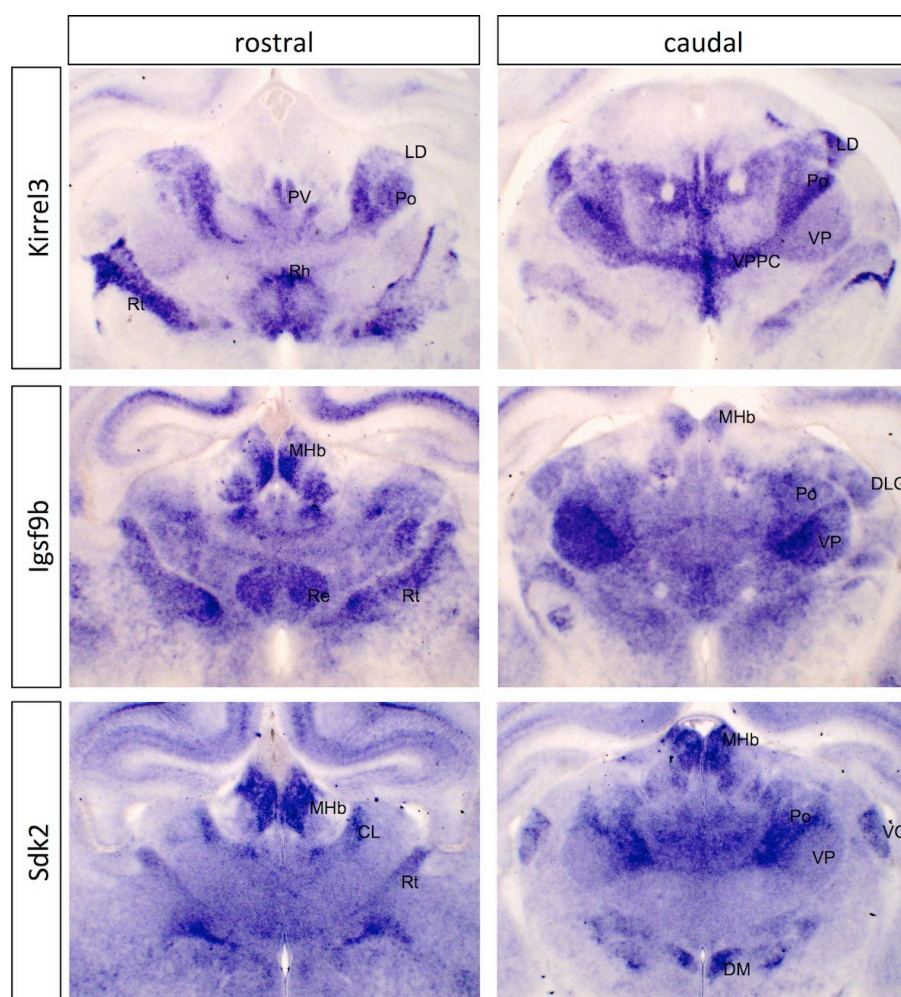

Figure 1. *In situ* hybridization patterns for immunoglobulin superfamily genes in dorsal thalamus at P0. Two coronal sections are shown for Kirrel3, Igsf9b and Sdk2, one rostral and one more caudal. The entire corresponding sections are shown in Figure S1, for context. CL, centrolateral nucleus; DLG, dorsolateral geniculate

nucleus; DM, dorosmedial hypothalamic nucleus; LD, laterodorsal nucleus; MHb, medial habenula; Po, posterior thalamic nuclear group; PV, paraventricular nucleus; Re, reuniens nucleus; Rh, rhomboid nucleus; Rt, reticular nucleus; VG, ventral geniculate nucleus; VP, ventral posterior nucleus; VPPC, ventral posterior nucleus parvicellular part.

Odz (Teneurin) genes: We found that all four members of the Odz/Teneurin family are expressed in selective fashion across the developing thalamus at P0 (Figure 2; Figure S2) and at E15 (Figures S3-S4). These genes are orthologues of the fly *odd oz/teneurin* genes *ten-a* and *ten-m,* which function in synaptic connectivity in the olfactory and neuromuscular systems [reviewed in 34]. In mammals, they have both homophilic and heterophilic interactions. Odz2 (Ten-m2) and Odz4 (Ten-m4) are expressed in similar, but not identical patterns, and Odz3 (Ten-m3) expression overlaps substantially with them. Odz1 (Ten-m1) is expressed in a broadly complementary pattern to the other three genes. These genes also show graded expression across cortex and striatum at E15 and P0 (Figures S2 and S4). Roles for Odz2 and Odz3 have been demonstrated in establishing the topography and segregation of ipsilateral and contralateral afferents in the dLGN and of thalamic afferents in visual cortex [reviewed in 35]. More recently, Odz3 has also been shown to control topography of thalamostriatal projections [13].
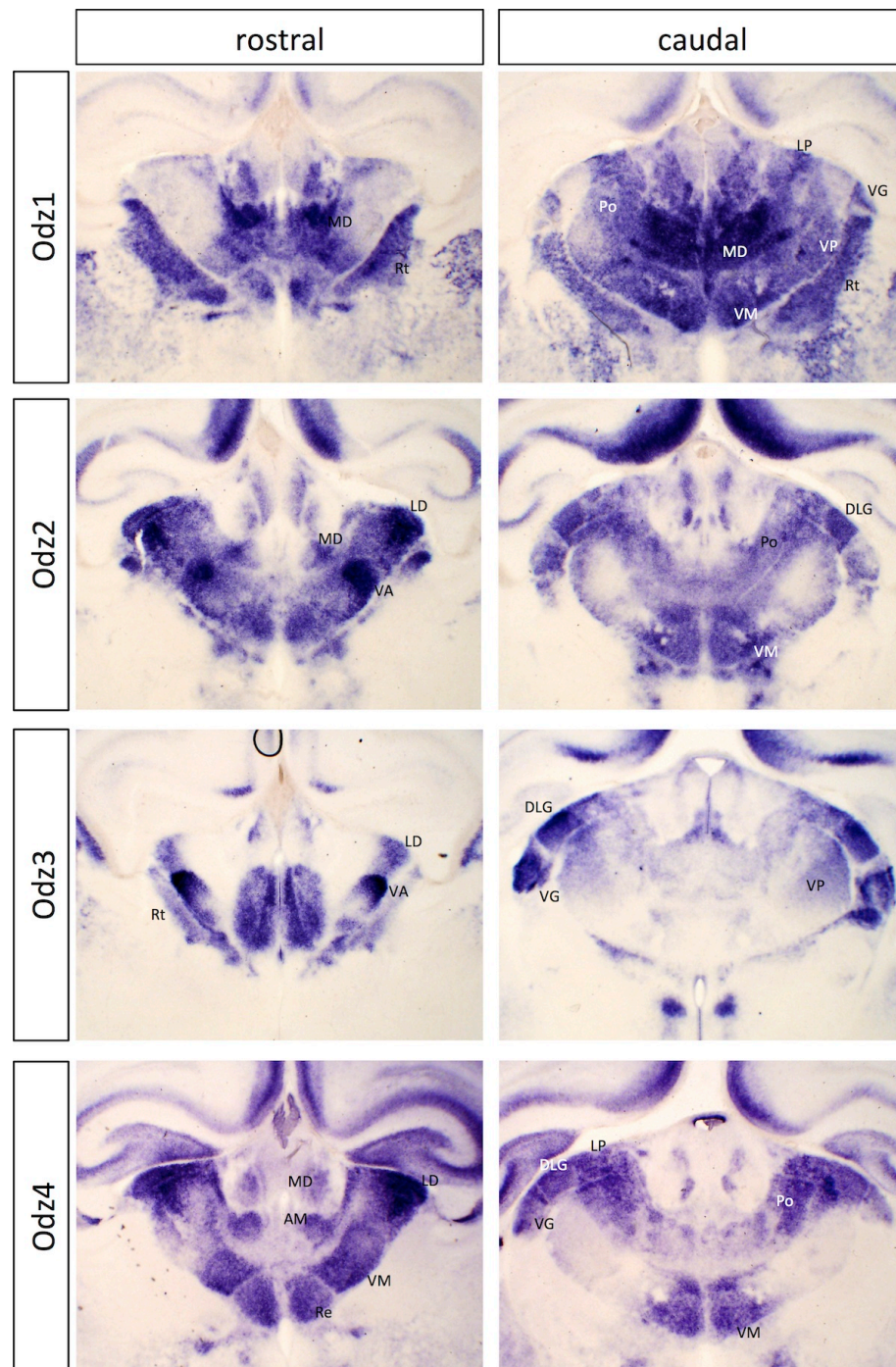
Figure 2. *In situ* hybridization patterns for Odz family genes in dorsal thalamus at P0. Two coronal sections are shown for Odz1, Odz2, Odz3 and Odz4, one rostral and one more caudal. The entire corresponding sections are shown in Figure S2, for context. AM, anteromedial nucleus; DLG, dorsolateral geniculate nucleus; LD, laterodorsal nucleus; MD, mediodorsal nucleus; Po, posterior thalamic nuclear group; Re, reuniens nucleus; Rt, reticular nucleus; VA, ventral anterior nucleus; VG, ventral geniculate nucleus; VM, ventromedial nucleus; VP, ventral posterior nucleus.

Neto genes: The neuropilin and tolloid-like genes Neto1 and Neto2, orthologues of the fly gene *neto*, also showed discrete expression in the thalamus, with Neto2 more widely expressed (Figure 3; Figure S5). In some parts of the thalamus, their expression was strikingly complementary, particularly in the ventrobasal complex, which expresses high levels of Neto1 but not Neto2, while many other nuclei show the opposite pattern. The Neto proteins function as auxiliary subunits for kainate receptors in mammals [reviewed in 36] and in glutamate receptor clustering at the neuromuscular junction in flies [37]. No role in axon guidance or synaptic connectivity has been demonstrated for these genes, but their early expression (Figure S6) and protein similarity to neuropilins would certainly be consistent with such a function.
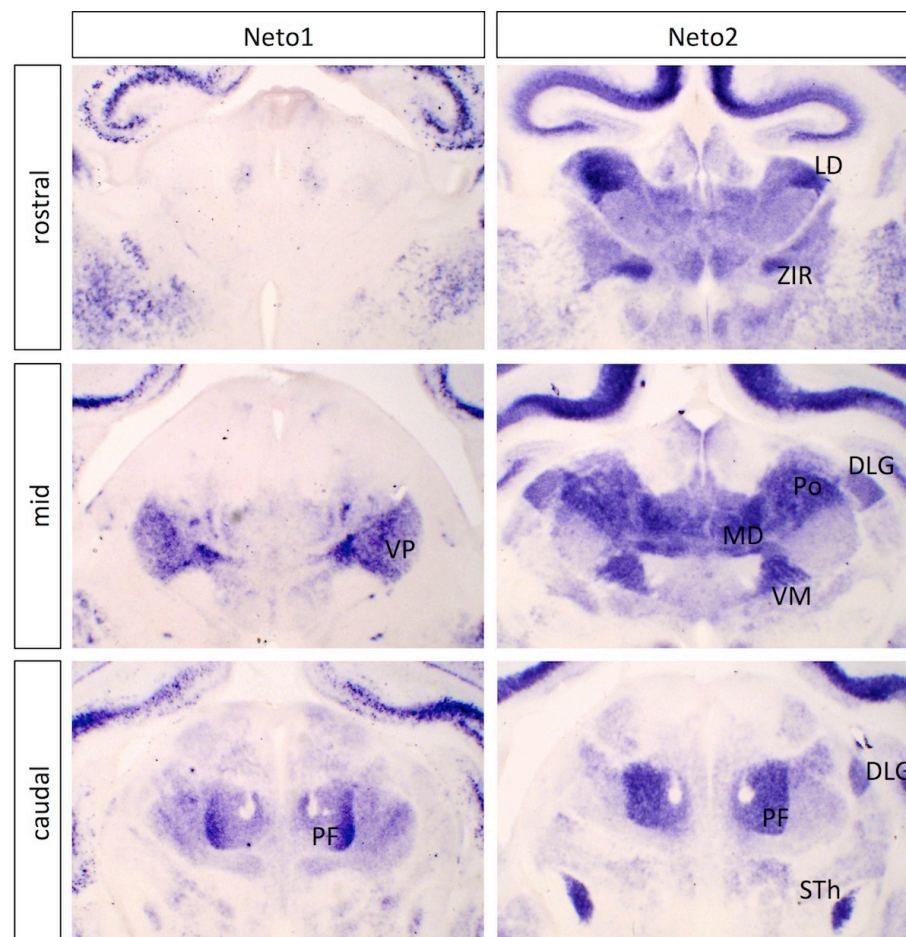

Figure 3. *In situ* hybridization patterns for Neto family genes in dorsal thalamus at P0. Three coronal sections are shown for Neto1 and Neto2, one rostral, one at an intermediate level (mid) and one more caudal. The entire corresponding sections are

shown in Figure S5, for context. Expression of the Neto genes is notably complementary in some places (compare mid sections) and overlapping in others (e.g., PF). DLG, dorsolateral geniculate nucleus; LD, laterodorsal nucleus; MD, mediodorsal nucleus; PF, parafscicular nucleus; Po, posterior thalamic nuclear group; STh, subthalamic nucleus; VM, ventromedial nucleus; VP, ventral posterior nucleus; ZIR, zona incerta rostral part.

Calsyntenins encode transmembrane proteins with two extracellular cadherin domains [38]. They have recently been implicated in synaptogenesis through interactions with neurexins [39]. As with previous reports, we find Clstn1 to be broadly (but not ubiquitously) expressed across the thalamus [38,39], while Clstn2 is expressed more selectively (Figure 4; see also Figure 17, below), in contrast to report by Hintsch *et al.* of only background expression. Clstn3 mRNA was not detectable.
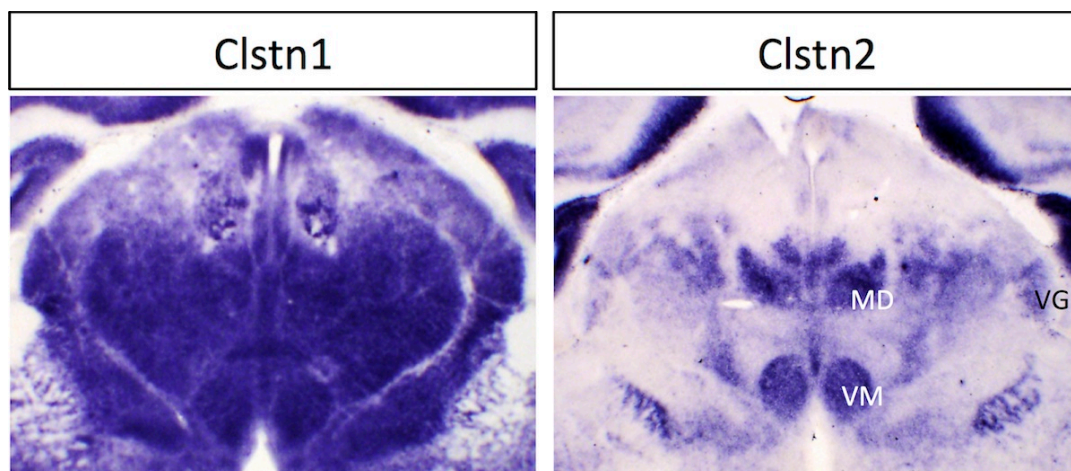
Figure 4. *In situ* hybridization patterns for Clstn family genes in dorsal thalamus at P0. One coronal section for Clstn1 and Clstn2 is shown. Additional sections from the devABA are shown in Figure 17. MD, mediodorsal nucleus; VG, ventral geniculate nucleus; VM, ventromedial nucleus.

Clustering analyses of developmental gene expression patterns in the thalamus

As a complementary strategy to identify candidate connectivity labels, we analysed data from the devABA, extending our search beyond genes conserved between vertebrates and invertebrates. The devABA provides qualitative and quantitative data

on expression of 2002 genes based on mRNA *in situ* hybridization. These genes were chosen based on functional relevance to brain development or disorders of the brain. The expression density values for each gene in the atla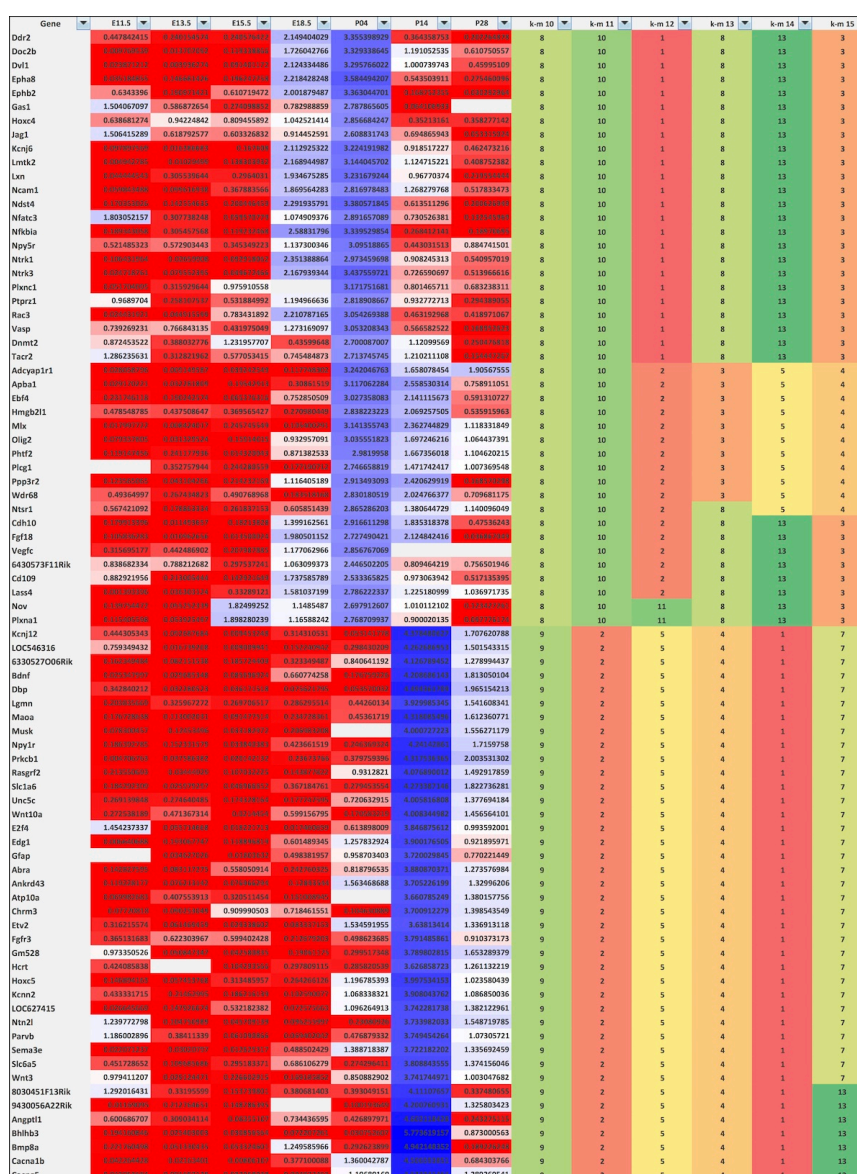s have been mapped to a three-dimensional digital template, with individual voxels attributed to different brain regions. We therefore extracted the expression density values for voxels attributed to the dorsal thalamus for 2002 genes at each of seven ages: embryonic day 11.5 (E11.5), E13.5, E15.5, E18.5, postnatal day 4 (P4), P14 and P28 (Table S2). These time-points are developmentally relevant to regional specification/patterning (E11.5), axon pathfinding (E13.5, E15.5 and E18.5), synaptogenesis (P4 and P14) and cortical plasticity (P28). Because we were more interested in comparing temporal profiles than relative levels across genes, we normalised these data for each gene by dividing by the average expression value for that gene across all ages (Table S3 Columns B-H).

In order to find genes with similar temporal expression profiles we used k-means clustering, which clusters elements of a matrix into a user-defined number of clusters (the k value). Since there is not necessarily a "correct" number of clusters, we performed k-means clustering with values of k from 6 to 18 and examined the flux in clustering outputs across these levels. As the input value of k increased, clusters tended to get subdivided as opposed to reshuffled, giving a roughly hierarchical organisation (Table S3 Columns I-U; Figure 5). We selected $k = 10$ for further data analysis, since the data segregated into clusters differentially peaking at all embryonic and postnatal time-points, while leaving large enough clusters for statistical analyses of enrichment (Figure 6). Lower k values gave poorer separation across the embryonic and early postnatal ages of highest interest to us, while higher values gave increasing refinement only at later postnatal stages (Figure S7).

Figure 5. Excerpt of *k*-means clustering analyses. Clustering of a small subset of genes is shown to illustrate the method. Gene names are shown in column 1. Columns 2 to 8 show normalised expression densities per gene per time-point. Heatmap's 3 colour scale of gene expression data: 0.2, red; 1, white; 5, blue. Columns 9 to 14 show cluster numbers assigned under $k = 10$ to $k = 15$ clustering, with clusters colour-coded to facilitate visualisation. Clusters are sorted at $k = 10$, followed by $k = 11$, followed by $k = 12$ and so on until $k = 15$. As the input value of k increases, clusters tend to get subdivided as opposed to reshuffled, giving a roughly hierarchical organisation. A clear difference in expression pattern between clusters 8 and 9 at $k = 10$ is evident, with cluster 8 showing peak expression at P4, and cluster 9 showing peak expression at P14. Within cluster 8, additional subclusters become apparent at $k = 12$ and above. The full clustering output for all 1996 genes is shown in Table S3.

13

k-mean 10	E11.5	E13.5	E15.5	E18.5	P04	P14	P28	# genes
cluster 7	4.087	0.733	0.318	0.503	0.430	0.478	0.374	131
cluster 6	2.067	0.665	0.465	0.791	0.785	1.423	0.820	268
cluster 5	0.932	2.706	0.674	0.686	0.560	0.946	0.447	106
cluster 2	0.769	0.966	1.548	1.047	1.196	0.769	0.703	163
cluster 4	0.553	0.378	0.426	2.779	1.286	1.003	0.550	157
cluster 0	0.317	0.318	0.421	0.965	1.628	2.069	1.256	384
cluster 8	0.403	0.284	0.303	0.791	3.467	1.070	0.662	192
cluster 9	0.432	0.249	0.226	0.482	0.491	4.237	0.839	214
cluster 3	0.271	0.213	0.190	0.329	0.650	2.949	2.374	253
cluster 1	0.450	0.294	0.225	0.494	0.566	1.175	3.699	128

Figure 6. Summary of expression profiles at k = 10. Normalised expression densities were averaged per cluster to see the trends of expression at k = 10. Clusters were organised chronologically with those showing early peaks of expression at the top and later peaks of expression at the bottom. Heatmap's 3 colour scale of gene expression data: 0.2, red; 1, white; 5, blue. Number of genes per cluster is shown in the rightmost column.

Gene annotation and enrichment analyses

We used two different approaches to annotate genes, based on known or predicted localisation of the protein product or on known biological function. First, we screened protein products for cellular localisation with an emphasis on the extracellular environment by assessing the outputs of several software tools available online (PSORTII, SignalP, Big-PI Predictor and TMHMM; Table S4), along with manual curation based on literature sources. The devABA dataset predominantly comprises proteins found in the nucleus (907 proteins), followed by 293 secreted proteins, 286 GPI-anchored or single-pass transmembrane proteins, 267 cytoplasmic (or in organelles) proteins and 240 multi-pass transmembrane proteins (Table S5 Column K). Three genes could not be annotated: C030002O17Rik, LOC433436 and mCG146432.

Second, we grouped genes by ascribed functions, using GO term annotations and systematic manual curation (Table S5 Column M). The dataset was divided into 6 groups: Group 1: axon guidance pathway and cell adhesion; Group 2: synapse; Group

3: receptor tyrosine kinases and their ligands, and patterning; Group 4: neurotransmission pathway (G-protein-coupled receptors, ion channels, gap junctions); Group 5: chromatin and transcription factor activity; Group 6: other (cytoskeleton, extracellular matrix, myelin, metabolic enzymes and signal transduction) and unannotated.

When a gene fell into more than one group, an order of priority was used (the order in which the groups are presented); for example, a gene encoding a receptor tyrosine kinase that is involved in axon guidance would appear in Group 1 and not in Group 3. Our dataset predominantly included genes involved with chromatin and transcription factor activity (677 genes in Group 5); followed by 360 genes expressing receptor tyrosine kinases and their ligands, and genes involved in patterning (Group 3); 286 genes involved in the axon guidance pathway and cell adhesion (Group 1); 270 in the neurotransmission pathway (Group 4); and, 91 genes involved with synapses (Group 2; Table S5). Group 6 contained 107 genes involved with or part of the cytoskeleton, extracellular matrix, myelin, signal transduction pathways or expressed metabolic enzymes. This group also included 205 unannotated genes to bring the total to 312 genes.

We hypothesized that different types of proteins would be enriched in clusters with peak expression at time-points relevant to specific developmental functions. In order to test this, we compared the observed numbers for each category in each cluster to the expected value by normal distribution.

A variety of trends emerged from these data (Figures 7 and 8). Chi square statistics on the whole dataset showed deviation from expected by random distribution at p-value < 0.001 for either localisation or function (Table 2A and B). Further statistics were done on individual categories to establish which ones were enriched in particular clusters. The Bonferroni correction was taken into account due to multiple testing, lowering the threshold to p-value ≤ 0.01 and p-value ≤ 0.008 for localisation and function, respectively. For localisation, the trends observed with secreted and cytoplasmic proteins did not reach statistical significance (p-value > 0.1). However,

the distribution of single-pass transmembrane and GPI-anchored proteins, multi-pass transmembrane proteins and nuclear proteins were significant at p-value < 0.0001, p-value < 0.001 and p-value < 0.001, respectively. For functional groups, all trends observed reached statistical significance, at p-value < 0.005 for Group 2, p-value < 0.001 for Group 1, and p-value < 0.0001 for Groups 3 to 6.

Table 2A Enrichment Analyses by Protein Localisation

Cluster	Chi Square	Degrees of Freedom	P-Value	Bonferroni Correction	Conclusion
All	122.694	36	< 0.001	No	Enrichment
Secreted	16.328	9	> 0.1	Yes	No Enrichment
Single-Pass and GPI	42.006	9	< 0.0001	Yes	Enrichment
Multi-Pass	31.815	9	< 0.001	Yes	Enrichment
Cytoplasm	13.559	9	> 0.1	Yes	No Enrichment
Nucleus	65.758	9	< 0.0001	Yes	Enrichment

Degrees of freedom: All, (columns # - 1) (row # - 1) = (5 - 1) (10 - 1) = 36; Per Localisation, (c - 1) (r - 1) = (10 - 1) (2 - 1) = 9. Bonferroni correction: $p \leq \alpha$ / # tests = 0.05 / 5 = 0.01

Table 2B Enrichment Analyses by Functional Group

Cluster	Chi Square	Degrees of Freedom	P-Value	Bonferroni Correction	Conclusion
All	206.160	45	< 0.001	No	Enrichment
Group 1	30.770	9	< 0.001	Yes	Enrichment
Group 2	26.554	9	< 0.005	Yes	Enrichment
Group 3	46.368	9	< 0.0001	Yes	Enrichment
Group 4	41.989	9	< 0.0001	Yes	Enrichment
Group 5	73.622	9	< 0.0001	Yes	Enrichment
Group 6	37.325	9	< 0.0001	Yes	Enrichment

Degrees of freedom: All, (columns # - 1) (row # - 1) = (6 - 1) (10 - 1) = 45; Per Group,(c - 1) (r -

1) = (10 - 1) (2 - 1) = 9. Bonferroni correction: $p \leq \alpha$ / # tests = 0.05 / 6 = 0.008

By plotting the relative enrichment for each category across clusters, ordered by time-point of highest expression, it was possible to discern what relationships were driving these deviations from a random distribution (Figures 7 and 8), and to assess their biological plausibility. For the localisation data, nuclear proteins were enriched in clusters peaking at early time-points, whereas single-pass transmembrane and GPI-anchored proteins were enriched in clusters peaking from E15.5 to P4, and multi-pass transmembrane proteins in clusters peaking from P4 to P28 (Figure 7). This nicely correlates with the extensive requirement of transcription factors during patterning and differentiation early on (localised in the nucleus), followed by axon guidance molecules expressed extracellularly during axon pathfinding, and finally synaptogenesis and neurotransmission postnatally (including many multi-pass neurotransmitter receptors and ion channels). These inferences were supported by enrichment of functional groups: chromatin and transcription factors early on, axon guidance and synapse molecules in later embryogenesis, and neurotransmission pathway genes at later postnatal stages (Figure 8).
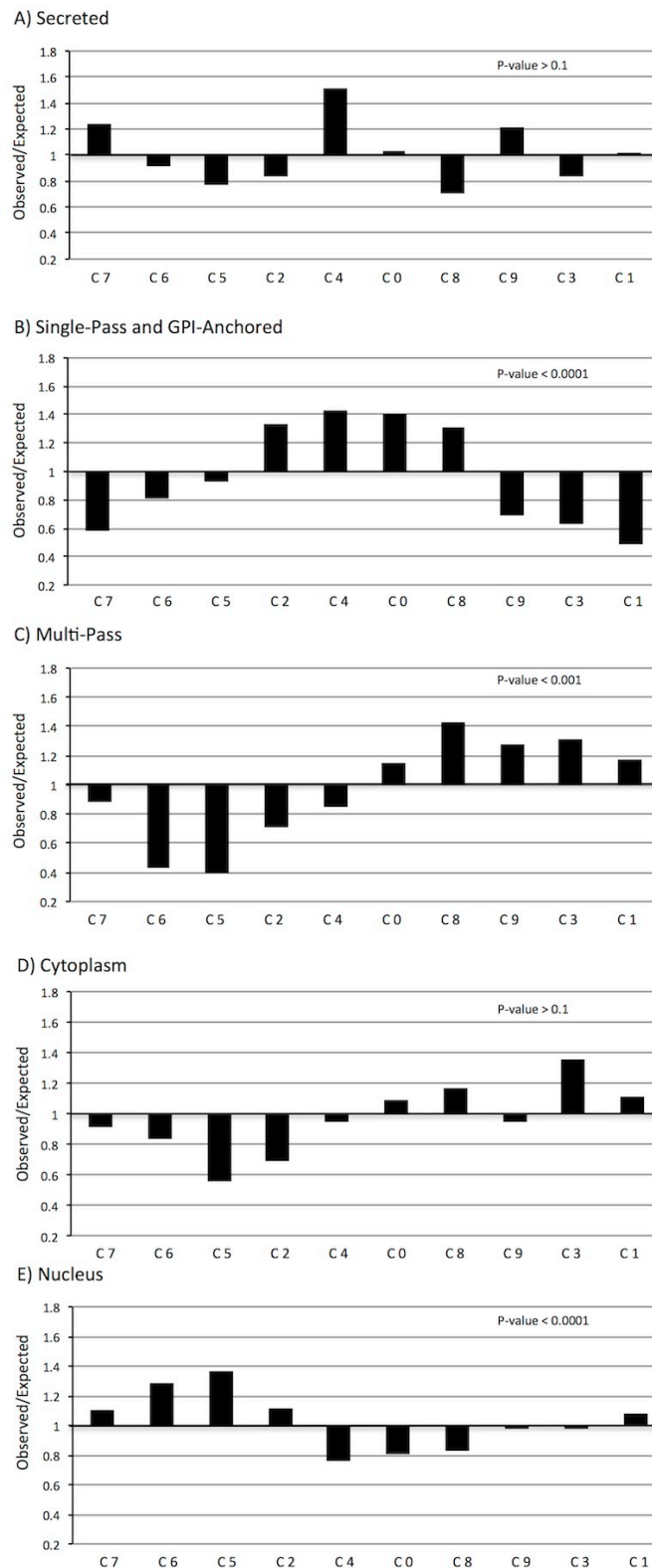
Figure 7. Enrichment of protein localisation categories across clusters. Relative enrichment across clusters is shown by plotting observed/expected numbers of proteins per cluster for each of five groups: secreted proteins (A), single-pass transmembrane or GPI-anchored proteins (B), multi-pass transmembrane proteins (C),

cytoplasmic proteins (D) and nuclear proteins (E). A value of 1 indicates observed data matches expected values, whereas a value below or above 1 indicates decreased or increased counts compared to expected values, respectively. The ten clusters (C0-C9) are organized according to age of peak expression, as in Figure 6. P-values from chi square analyses are shown in upper right corner of each graph. Genes encoding nuclear proteins are enriched in clusters defined by strong early embryonic expression, single-pass transmembrane and GPI-linked protein-encoding genes are enriched in clusters expressed at mid-embryonic stages and multi-pass transmembrane proteins are enriched in clusters expressed at postnatal stages.
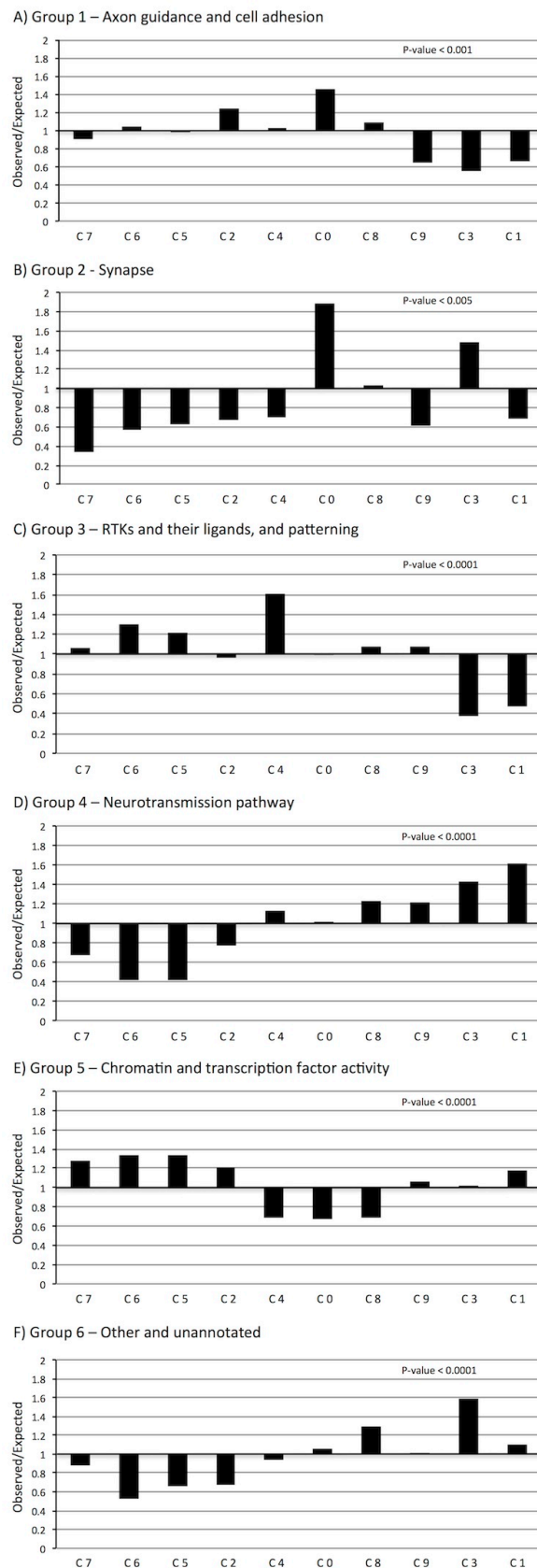
Figure 8. Enrichment of functional groups across clusters. Relative enrichment across clusters is shown by plotting observed/expected numbers of proteins per cluster

for each of six groups: Group 1 - axon guidance pathway and cell adhesion (A); Group 2 - synapse (B); Group 3 - receptor tyrosine kinases and their ligands, and patterning (C); Group 4 - neurotransmission pathway (GPCRs, ion channels, gap junctions; D); Group 5 - chromatin and transcription factor activity (E); and, Group 6 - other (cytoskeleton, extracellular matrix, myelin, metabolic enzymes and signal transduction) and unannotated (F). A value of 1 indicates observed data matches expected values, whereas a value below or above 1 indicates decreased or increased counts compared to expected values, respectively. Clusters 0-9 are organized chronologically. P-values from chi square analyses are shown in upper right corner of each graph. All groups showed statistically significant deviation from expected distributions.

Taken together, these protein localisation and functional enrichment analyses provide a powerful and independent validation of our clustering analyses, indicating that the clusters we have identified with $k = 10$ track valid biological trends. They further suggest that clusters enriched for known molecules involved in axon guidance or synaptic connectivity may also contain novel molecules with these functions. In order to specifically search for such genes, we focused on clusters 2, 4, 0 and 8, which show enrichment for single-pass and GPI-linked proteins and for proteins involved in axon guidance or synaptogenesis. These have peak expression at E15.5 (cluster 2), E18.5 (cluster 4) and P4 (clusters 0 and 8).

Identifying candidate connectivity labels

Clusters 2, 4, 0 and 8 had a total count of 896 genes of which 426 encode extracellular proteins. We screened through them by visual inspection of the devABA expression data at their peak expression to identify those with selective or differential expression across the thalamus. Genes showing uniform expression were not considered further. We generated a list of 215 genes that potentially encode specific connectivity information (Table S6). This dataset includes numerous genes already implicated in thalamocortical connectivity (such as Chl1, DCC, L1CAM, NCAM1, Ntn1, Robo1 and Robo2), which we did not characterize further. However, where previously implicated genes were part of larger families (e.g., cadherins, Ephrins and Eph-

receptors, semaphorins), they were included for comparison with other members. Some genes had previously demonstrated roles in axon guidance or synapse formation in other contexts and were of high interest to us, as were those encoding surface proteins not previously implicated in these processes. However, we did not pursue any of the large number of neurotransmitter receptor and ion channel genes, given a lower prior probability of involvement as direct axon guidance or synaptic connectivity labels.

In order to present a summary of the expression patterns of these genes and to enable comparison across genes, we extracted images for each gene from the devABA database of a consistent lateral and/or a more medial or parasagittal section at E18.5 (Figure S8 and Figures 9-20). As the borders between nuclei are still developing at E18.5, it is not possible to definitively match expression patterns to specific nuclei at this stage. These figures are not intended to be comprehensive, as the full dataset can be viewed on the devABA website. They do, however, allow a survey of the general trends of expression patterns across the thalamus and a comparison of multiple members of specific gene families, while also highlighting numerous individual genes with strikingly selective patterns.

In general, the expression patterns were extremely diverse – in fact, we did not detect any two patterns that were identical across these genes. Some genes showed highly selective expression patterns, on in some developing nuclei and off in others. Others showed differential expression, higher in some nuclei than others, with substantial variation in extent of expression across the thalamus. Within developing nuclei, some genes appeared to be expressed at uniform levels, while others showed an uneven or graded distribution. Below, we consider the expression patterns of 73 genes organised by protein family or general function (Table 1 Columns B-C).

Ephrins and Eph-receptors: Twelve members of the Eph and Ephrin gene families show differential or selective expression across the developing thalamus (Figure 9). Multiple members of the EphA and -B receptor and Ephrin-A and -B families have been implicated in diverse aspects of thalamic connectivity. The most extensively investigated have been EphA4, EphA7, EphrinA2 (Efna2) and EphrinA5 (Efna5). Roles for these genes have been demonstrated in parcellation of thalamic nuclei [19],

target selection and topography of retinal axons into the thalamus [40], topographic guidance of thalamic axons through the ventral telencephalon [41], areal and layer-specific targeting and topography [reviewed in 42], reciprocal corticothalamic axon guidance and target selection [43,44] and even influences on cortical progenitor proliferation and differentiation dynamics [45]. EphB1 and -B2 have also recently been shown to mediate thalamic axon guidance through the ventral telencephalon [46]. The selective expression of several other members in the thalamus, including EphA3, -A5, -A6 and -A8 has also previously been noted [20,47]. The ABA data confirm these findings and allow a direct comparison of expression patterns across members of these families. In addition, we find that EphA1, EphA10, EphB6 and Efnb3 are also differentially expressed across the developing thalamus.

Figure 9. Ephrin and Eph-receptor expression in the thalamus at E18.5. *In situ* hybridization data from sagittal sections of the thalamus at E18.5 obtained from the devABA revealed differential patterns of expression of Eph-receptor (Eph) A1, A3, A4, A6, A8, A10, B1, B2, B6, and Ephrin (Efn) A2, A5 and B3. See text for details. A representative and equivalent section of medial thalamus is shown for each gene (see Figure S7 for details on how the sections were selected).

Cadherins: We report differential or selective expression of twelve members of the cadherin family across the developing thalamus (Figure 10). A number of cadherins have been implicated previously in thalamic connectivity. Takeichi and colleagues showed differential expression of Cdh6, Cdh8 and Cdh11 in both cortex and thalamus and used dye-tracing to demonstrate that Cdh6- or Cdh8-positive cortical areas are specifically connected with Cdh6- or Cdh8-positive thalamic nuclei, respectively [48]. Similarly, but for afferent connections to the thalamus, Cdh6 has been shown to act as a homophilic targeting molecule for a subset of retinal ganglion cells which project to nuclei mediating non-image-forming visual functions [10]. N-Cadherin (Cdh2) is required for proper termination of thalamic axons in layer 4 of the cortex [49], and Cdh2 and Cdh8 differentially label thalamic terminations in somatosensory cortex barrel center and septal compartments, respectively [50]. Differential expression across thalamic nuclei of the remaining cadherins in this group, Cdh4, 7, 9, 10, 12, 13, 24 and the atypical cadherin Fat3, has not, to our knowledge, been previously reported and greatly expands the possible involvement of this gene family in specifying thalamic connectivity. The diversity of expression patterns is particularly noteworthy; even in this single parasagittal section, no two members of this family show an identical pattern. Given the known potential for heteromeric complex formation between cadherins, these overlapping patterns generate a very large potential combinatorial code.

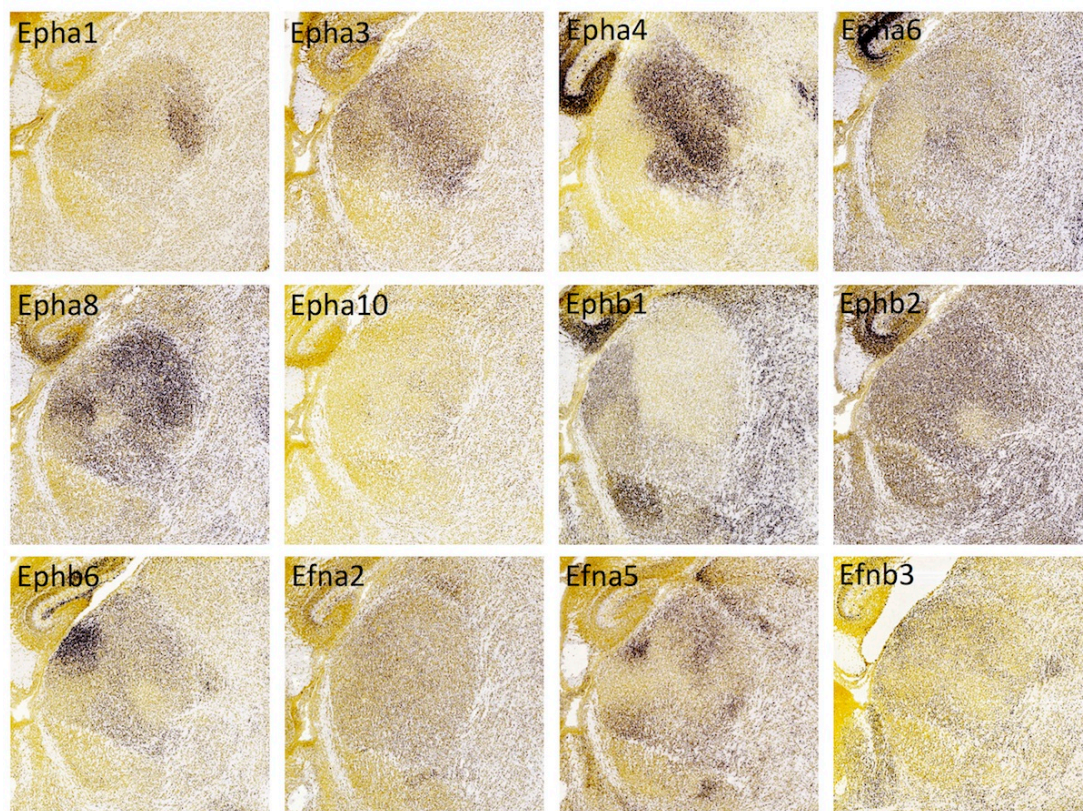**Figure 10. Cadherin expression in the thalamus at E18.5**. *In situ* hybridization data from sagittal sections of the thalamus at E18.5 obtained from the devABA revealed differential patterns of expression of Cadherin (Cdh) 2, 4, 6, 7, 8, 9, 10, 11, 12, 13, 24 and Fat3. See text for details. A representative and equivalent section of medial thalamus is shown for each cadherin (see Figure S7 for details on how the sections were selected).

Protocadherins: A number of members of the unclustered protocadherin family show differential expression across the thalamus at E18.5, including Pcdh1, 10, 11X, 19 and 21 (Figure 11). We note that Pcdh18 also shows differential thalamic expression at this stage (not shown) but is not included in our clusters of interest due to an earlier time-point of peak expression. These results replicate previous findings reported by Kim et al (2007), who also noted differential expression across thalamus of additional members of this family that are not in the devABA dataset (Pcdh7, 8, 9, 15 and 17) [51]. The functional importance of these genes in thalamic connectivity has not been directly tested. Though defects in thalamocortical axon projections occur in Pcdh10$^{-/-}$ mice, they have been attributed to functions in patterning of the ventral telencephalon

and not to activity in thalamic axons themselves [52]. Many of the protocadherins are also expressed in selective patterns across cortical areas and layers [51], suggesting a possible interaction code between interconnected regions.
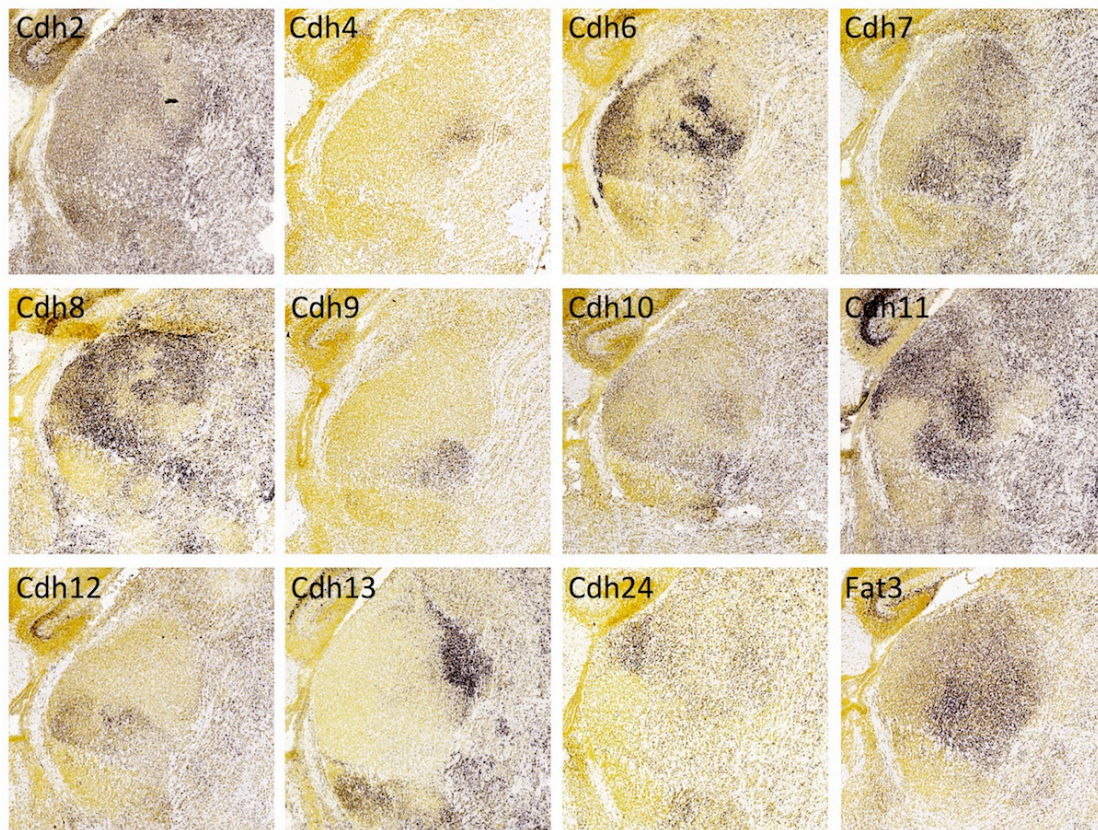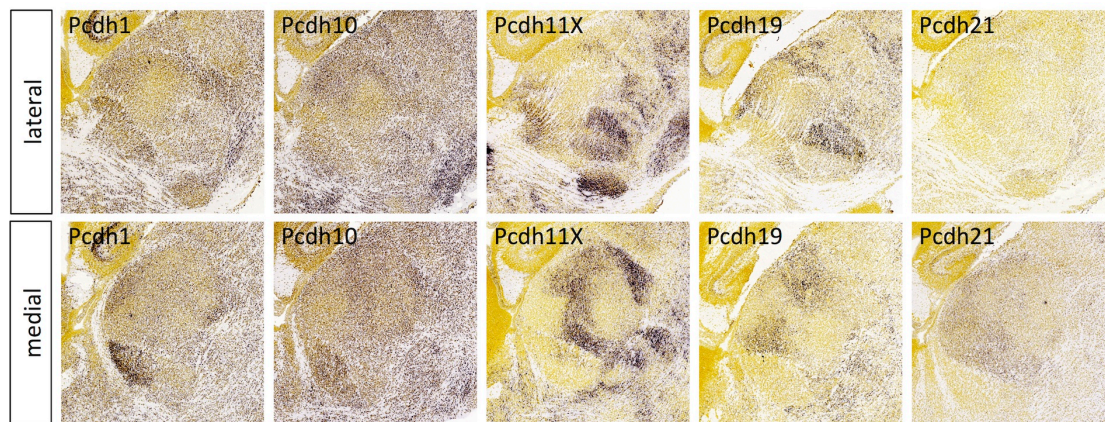

Figure 11. Protocadherin expression in the thalamus at E18.5. *In situ* hybridization data from sagittal sections of the thalamus at E18.5 obtained from the devABA revealed differential patterns of expression of Protocadherin (Pcdh) 1, 10, 11X, 19 and 21. See text for details. Two representative and equivalent sections are shown for each protocadherin, one lateral and the other medial (see Figure S7 for details on how the sections were selected).

Semaphorins and Plexins: Among the members of the semaphorin and plexin gene families in the devABA dataset, the ones showing the most selective or differential expression across the thalamus are: Sema3F, Sema6A, Sema7A, PlxnA1, PlxnA2 and PlxnC1 (Figures 12, 13). PlxnB1 and Sema6D are also expressed in developing thalamus, but weakly and more uniformly (not shown). Sema3F has been previously implicated in thalamic axon guidance through the ventral telencephalon, and Sema7A in regulating thalamocortical axon branching in layer 4 of cortex [53,54]. However, these functions depend on expression of the semaphorins in these other regions, not in the thalamus itself, where differential expression has not been previously reported. Null mutants of *Sema6A* show misrouted thalamocortical axons [55], specifically from the dLGN [56]; this function may depend on expression of Sema6A in the thalamus itself or in the ventral telencephalon. No roles in thalamic axon guidance or connectivity have yet been described for PlxnA1, PlxnA2 or PlxnC1. The expression of Sema7A and its receptor PlxnC1 are notably complementary, while Sema6A and

its binding partner PlxnA2 are expressed in an overlapping fashion, with Sema6A more widespread and PlxnA2 restricted to medial regions. Neuropilins also act as co-receptors for secreted semaphorins and have been implicated in thalamic axon guidance in response to Sema3F, signaling through Neuropilin-2-NrCAM co-receptor complexes [57] or Sema3A, acting through Neuropilin-1-CHL1 co-receptors [58]. Npn2 is not in the ABA developmental dataset but Npn1 is found in cluster 4 and is differentially expressed across thalamic nuclei, as previously described (not shown).

Figure 12. Semaphorin expression in the thalamus at E18.5. *In situ* hybridization data from sagittal sections of the thalamus at E18.5 obtained from the devABA revealed differential patterns of expression of semaphorins Sema3F, Sema6A and Sema7A. See text for details. Two representative and equivalent sections are shown for each semaphorin, one lateral and the other medial (see Figure S7 for details on how the sections were selected).
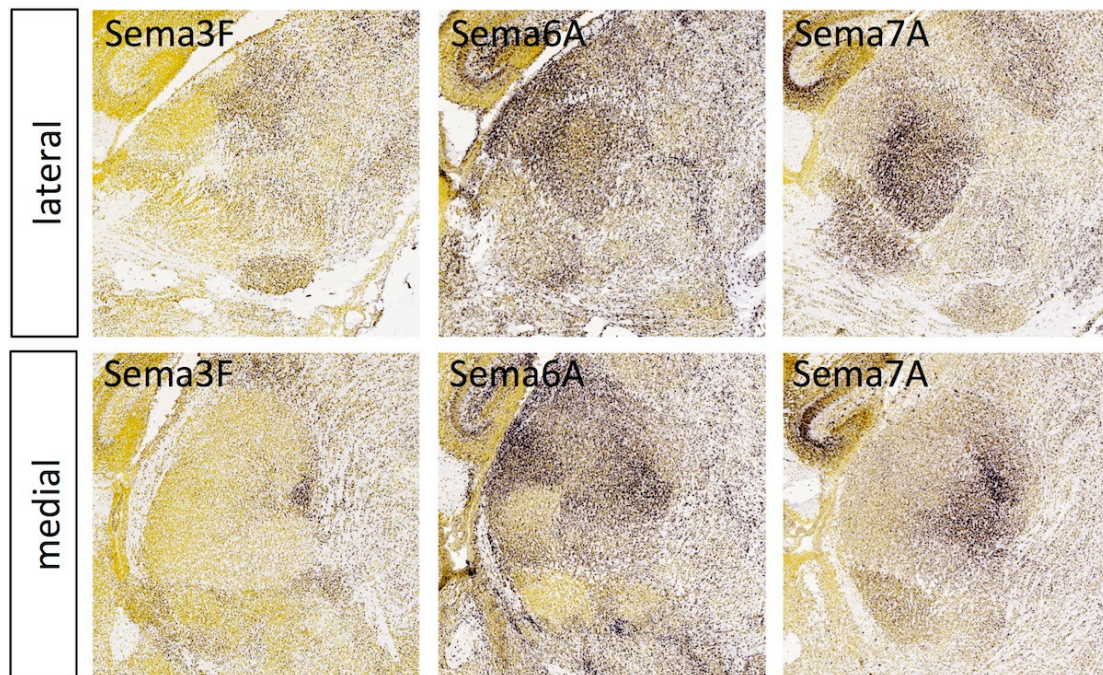
Figure 13. Plexin expression in the thalamus at E18.5. *In situ* hybridization data from sagittal sections of the thalamus at E18.5 obtained from the devABA revealed differential patterns of expression of plexins PlxnA1, PlxnA2 and PlxnC1. See text for details. Two representative and equivalent sections are shown for each plexin, one lateral and the other medial (see Figure S7 for details on how the sections were selected).

Receptor tyrosine kinases and phosphatases: In addition to the Eph family of receptor tyrosine kinases, we identified a number of other members of this superfamily and one receptor tyrosine phosphatase showing quite selective expression in the thalamus (Figure 14). Kinases Flt3, Kit (cKit) and Ret have well-documented functions in the immune system and in various cancers, while their functions in the nervous system have been explored to varying extents. No functions for Flt3 in the nervous system have been reported. Kit has been shown to have effects on cortical neuronal migration and axonal extension [59], but has not been functionally implicated in thalamic development, though Kit-ligand (Steel) is also highly expressed in developing thalamus [20]. Ret has multiple, well-described roles in development of the enteric nervous system, as a receptor for GDNF, and mutations in this gene are an important cause of Hirschsprung disease. While expression in

thalamus has been noted before [60], this has not been described in detail nor has a functional role in thalamic development been established. The neurotrophin receptors Ntrk2 (TrkB) and Ntrk3 (TrkC) have been studied in this context, with a demonstration that TrkB in particular is required for normal segregation of thalamic afferents in barrel cortex [61]. Ptpru (also called RPTP-lambda or PTP-RO) is closely related to RPTP-kappa and RPTP-mu and similarly mediates homotypic adhesion [62]. Ptpru has also been shown to associate with c-Kit and to negatively regulate its signaling [63]; however, the expression of these genes in thalamus appears largely non-overlapping (Figure 14). There are no published reports of Ptpru function in the nervous system, though many other members of the RPTP family play important roles in axon guidance and synaptic connectivity [64].
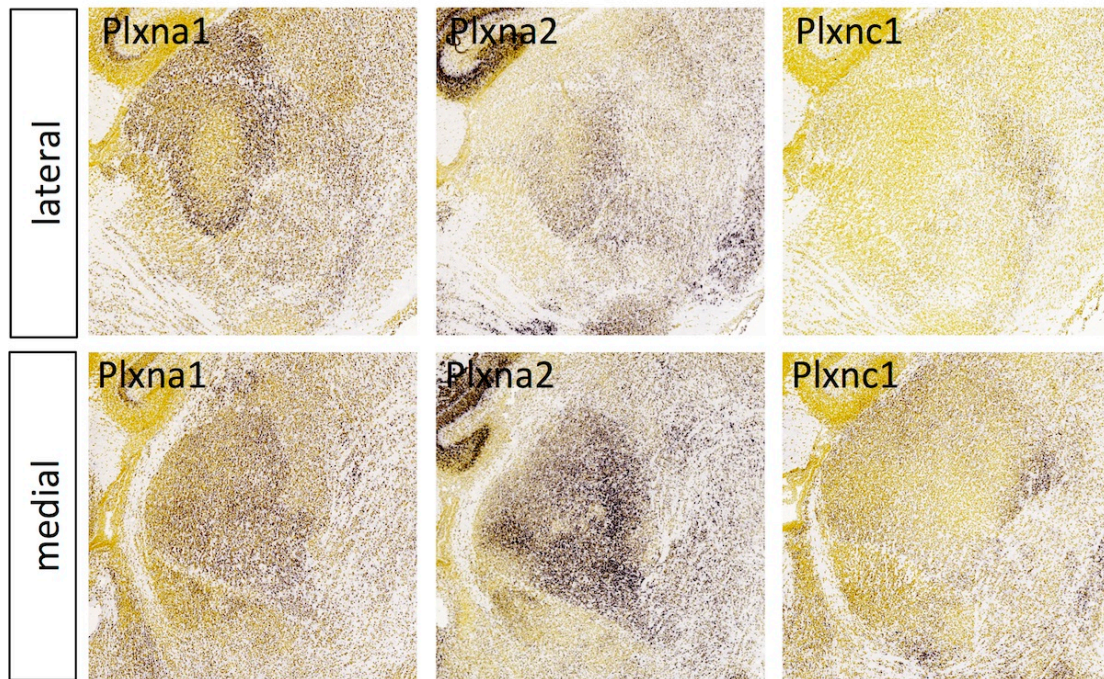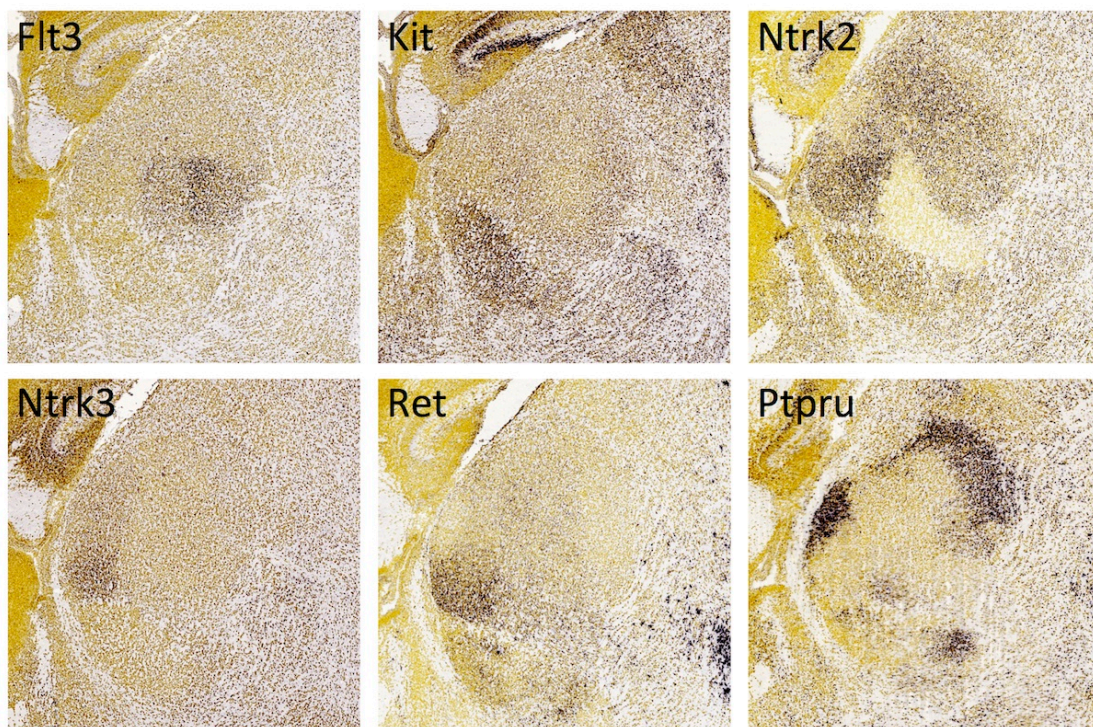

Figure 14. Expression of receptor tyrosine kinases and phosphatases in the thalamus at E18.5. *In situ* hybridization data from sagittal sections of the thalamus at E18.5 obtained from the devABA revealed differential patterns of expression of Flt3, Kit, Ntrk2, Ntrk3, Ret and Ptpru. See text for details. A representative and equivalent section of medial thalamus is shown for each gene (see Figure S7 for details on how the sections were selected).

Immunoglobulin superfamily. Multiple Ig superfamily molecules have been implicated in axon guidance or neuronal connectivity in the developing thalamus, including Robo proteins, DCC, various cell adhesion molecules (NCAM, L1, CHL1…) and others [reviewed in 9]. Here, we highlight the selective expression of five Ig superfamily members not previously implicated in thalamocortical development (Figure 15). Alcam (also known as CD166, Neurolin, or DM-GRASP/SC1/BEN) plays a role in guidance and fasciculation of motor and retinal axons [65], and in topographic mapping of retinal axons across the superior colliculus, possibly through interaction with adhesion molecule L1 on retinal ganglion cells [66]. Cadm1 (better known as SynCAM) is involved in synapse organisation [67] and axon guidance [68]. Cd47 is a multi-pass transmembrane molecule with a single extracellular Ig domain. It acts as a receptor for presynaptic organising molecule SIRPalpha [69] and has been implicated in various processes in cerebellar development [70]. Cntn6 (NB3) is a member of the contactin family of adhesion molecules, which have widespread roles in neural development [71]. *Cntn6* null mutants have defects in synapse formation in cerebellum [72] and hippocampus [73]. Our results replicate previous reports of expression of Cntn6 in anterior thalamic nuclei [74,75]. Mdga1 is expressed in a layer- and area-specific manner in the cortex [76] and mutation of the gene leads to a delay in cortical neuronal migration [77]. More recently, Mdga1 has also been found to interact with neuroligin-2 to negatively regulate inhibitory synapse formation [78,79].

Figure 15. Expression of immunoglobulin superfamily molecules in the thalamus at E18.5. *In situ* hybridization data from sagittal sections of the thalamus at E18.5 obtained from the devABA revealed differential patterns of expression of Alcam, Cadm1, Cd47, Cntn6 and Mdga1. See text for details. Two representative and

equivalent sections are shown for each gene, one lateral and the other medial (see Figure S7 for details on how the sections were selected).

Leucine-rich repeat superfamily: Many members of the extracellular leucine-rich repeat (LRR) superfamily have been implicated in axon guidance, synaptic target selection and other aspects of neural development [23,80]. LRR superfamily genes are highly under-represented in the devABA dataset, with no members of the Lrfn, Slitrk, NGL, Elfn or LINGO subfamilies and only one representative of each of the Lrrtm, Lrrc, Lrrn and Amigo subfamilies. Amongst the LRR genes that are represented, we identified three with selective expression in developing thalamus: Lgi2, Lrrn3 and Rtn4rl1 (Figure 16). Lgi2 encodes a secreted LRR protein of the Lgi subfamily that have been implicated in synaptic development and epilepsy [81]. Lgi1 regulates synapse formation and function through interactions with postsynaptic Adam22 and presynaptic Adam23 proteins [82,83]. Lgi2 likely has similar functions as it also interacts with these Adam proteins and mutations in the gene cause epilepsy in dogs [84]. Lrrn3 (or Nlrr3) is a member of the LRR-Ig-FN3 group [23]. Roles for Lrrn1 and Lrrn2 have been described in hindbrain development in chick [85,86]; otherwise this family remains poorly characterised. Selective expression in dorsal thalamus has been commented on previously, in particular its complementarity to the expression pattern of Slitrk6 [87]. Rtn4rl1 is also known as Nogo-receptor-3, NgR3. NgR's were first isolated as receptors for the inhibitory myelin protein Nogo-66, but also function as receptors for chondroitin proteoglycans [88] and have been implicated in synaptic development in hippocampus [89].
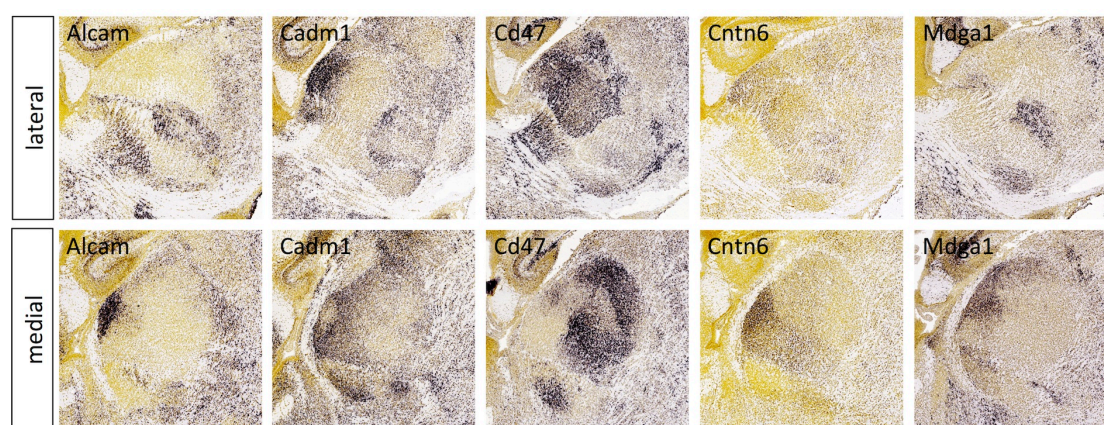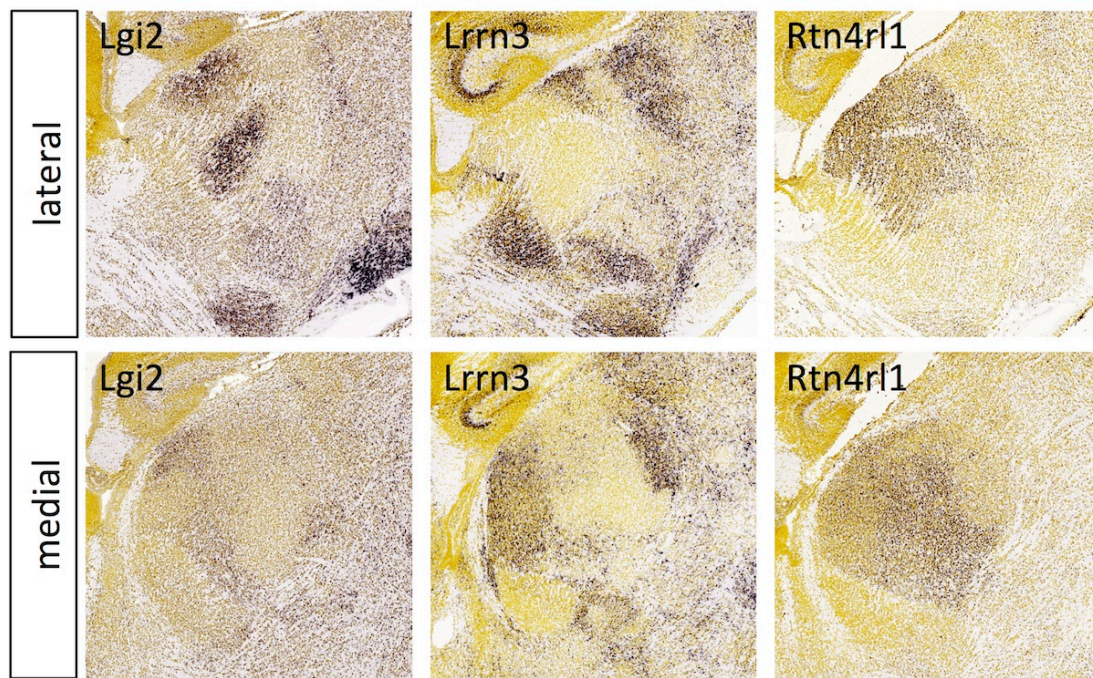
Figure 16. Expression of leucine-rich repeat superfamily members in the thalamus at E18.5. *In situ* hybridization data from sagittal sections of the thalamus at E18.5 obtained from the devABA revealed differential patterns of expression of Lgi2, Lrrn3 and Rtn4rl1. See text for details. Two representative and equivalent sections are shown for each gene, one lateral and the other medial (see Figure S7 for details on how the sections were selected).

Small gene families: A number of small gene families were identified with compelling expression patterns in developing thalamus (Figures 17, 18).

Calsyntenins: The devABA data replicate our own *in situ* hybridization results (Figure 17), with widespread, but not ubiquitous expression of Clstn1 and much more selective expression of Clstn2.

Cerebellins encode secreted molecules with synaptogenic activity. They interact with multiple partners including the glutamate receptor subunits GluRδ2 and neurexins [90] as well as the netrin receptor DCC [91]. We find strong and selective expression of Cbln2 and Cbln4 across thalamic nuclei (Figure 17). Their differential expression across the brain has been documented previously, including in the developing thalamus [92]. By contrast, Cbln3 is not expressed outside the cerebellum. Cbln1 is not in the devABA dataset but is also expressed in subsets of thalamic nuclei [92].
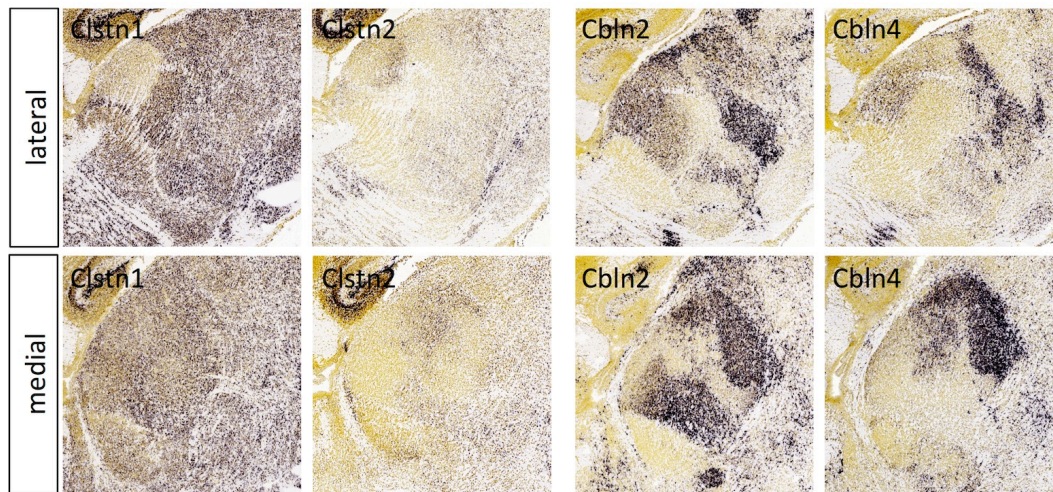
Figure 17. Calsyntenin and Cerebellin expression in the thalamus at E18.5. *In situ* hybridization data from sagittal sections of the thalamus at E18.5 obtained from the devABA revealed differential patterns of expression of Calsyntenin (Clstn) 1 and 2, and Cerebellin (Cbln) 2 and 4. See text for details. Two representative and equivalent sections are shown for each gene, one lateral and the other medial (see Figure S7 for details on how the sections were selected).

Netrin-Gs are GPI-anchored proteins related to netrins. They act as synaptic cell adhesion molecules via interactions with the transmembrane leucine-rich repeat proteins Netrin-G-ligand (NGL) -1 and -2 [93]. The devABA data match previous reports of broad Ntng1 expression across the dorsal thalamus and Ntgn2 expression restricted to the habenula (Figure 18) [94]. Ntng1 expression is not uniform, however; it displays significantly higher expression in some thalamic nuclei than others.
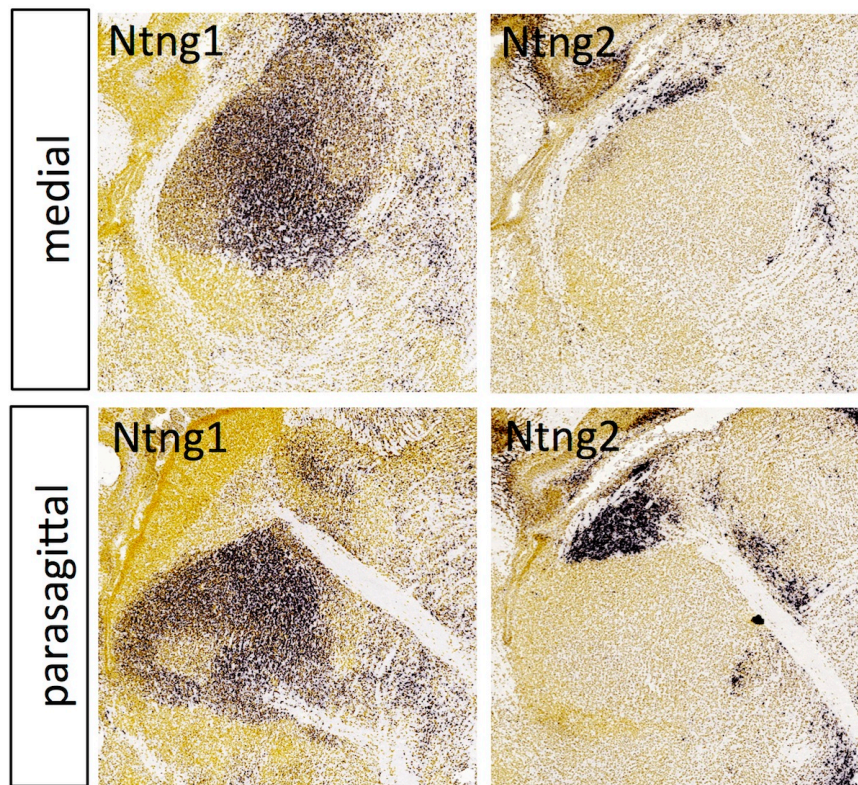
Figure 18. Netrin-G expression in the thalamus at E18.5. *In situ* hybridization data from sagittal sections of the thalamus at E18.5 obtained from the devABA revealed differential patterns of expression of Netrin-G (Ntng) 1 and 2. See text for details. Two representative and equivalent sections are shown for each netrin-G, one medial and the other adjacent medially (see Figure S7 for details on how the sections were selected).

Miscellaneous genes: In addition to the gene families described above, there is a large set of miscellaneous genes encoding surface or secreted proteins with differential expression patterns (Figure 19). These include: Astn2, Cntnap4, Dlk1, Dner, Fzd7, Gpc3, Lrp8, Lypd1 (also known as Lynx2), Odz3 and Trp53i11 (also known as Tp53i11 or PIG11). These display varying degrees of specificity. ,For example, Fzd7 is highly restricted to what may become a single thalamic nucleus, while Gpc3, is expressed in a small number of scattered cells throughout the thalamus (and strongly in the reticular nucleus, not shown). Astn2, Cntnap4, Dlk1, Lypd1 all show quite selective expression in some developing nuclei and not others, while Dner, Lrp8 and Trp53i11 are expressed more widely, but still differentially. The Odz3

expression pattern documented in the devABA mirrored that seen in our own *in situ* hybridization experiments extremely faithfully (Figure S9).
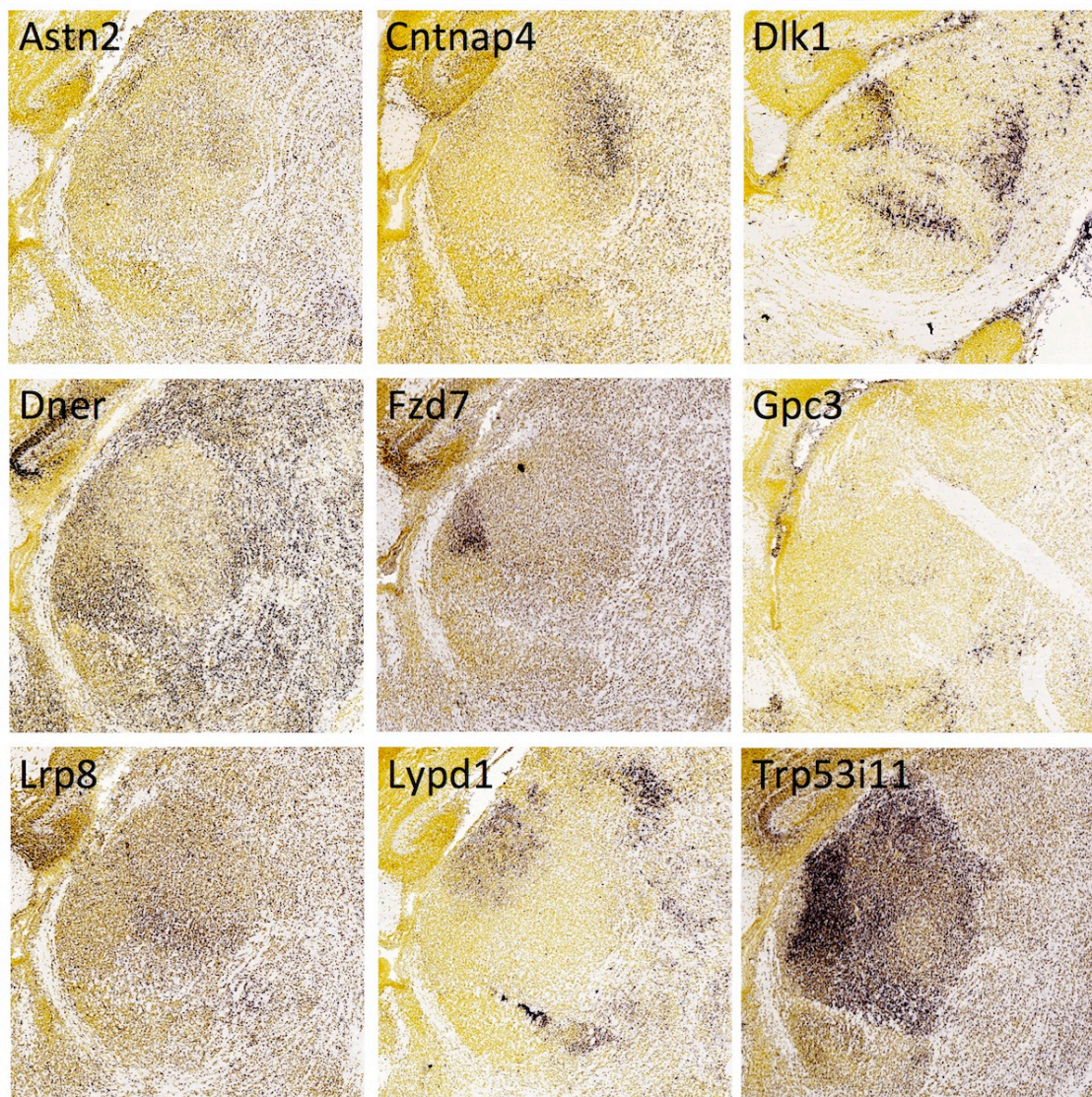
Figure 19. Expression of genes encoding miscellaneous surface or secreted molecules in the thalamus at E18.5. *In situ* hybridization data from sagittal sections of the thalamus at E18.5 obtained from the devABA revealed differential patterns of expression of Astrotactin2 (Astn2), Cntnap4, Delta-like-1 homolog (Dlk1), Delta/Notch-Like EGF Repeat Containing (Dner), Frizzled7 (Fzd7), Glypican-3 (Gpc3), Low-density lipoprotein receptor 8 (Lrp8), LY6/PLAUR Domain Containing 6 (Lypd1) and Trp53i11. See text for details. A representative and equivalent section is shown for each molecule (see Figure S7 for details on how the sections were selected). Dlk1 expression is shown laterally, while the other gene expression patterns are shown medially with Gpc3 more medial than the rest (parasagittal).

Lrp8 (low-density lipoprotein receptor 8) has been previously directly implicated in specifying thalamic connectivity. Mutation of *Lrp8*, along with the related gene *Vldlr (very low-density lipoprotein receptor)*, alters retinogeniculate innervation, through mechanisms that appear independent of the function of these proteins as Reelin receptors [11]. Mutation of *Cntnap4* affects synaptic outputs from GABAergic interneurons and ventral tegmental area dopaminergic neurons [95]. Gpc3 (Glypican-3) has been shown to bind to the synaptogenic molecule LRRTM4 [96] and, by analogy, with its close relative Gpc4, may actively specify synapse formation. Fzd7 (Frizzled-7) has not been directly implicated in synaptic connectivity but many other members of the Frizzled family have been [97], including Fzd5, which is required for the synapse-organising activity of Wnt7a [98]. Fzd5 is extremely selectively expressed in the parafascicular nucleus of thalamus [99], which we also find here (data not shown).

A variety of other functions in neural development or function have been shown for Astn2 (Astrotactin-2) [100], Dlk1 (Delta-like-1 homolog) [101], Dner (Delta/Notch-Like EGF Repeat Containing) [102] and Lypd1/Lynx2 (LY6/PLAUR Domain Containing 1) [103] but these genes have not been directly implicated in axon guidance or synaptic connectivity. The functions of Trp53i11 remain largely unknown.

Growth factors and receptors. A variety of growth factors and receptors show selective or differential expression across the late embryonic thalamus (Figure 20). These include: Bmp3 (Bone morphogenetic protein-3), Igfbp5 (Insulin-like growth factor binding protein 5), Inhba (Inhibin/Activin, beta A subunit), Gfra1 and Gfra2 (GDNF Family Receptor Alpha 1 and 2), Nrn (Neuritin or Neuritin-1/Nrn1), Tgfb2 (Transforming growth factor, beta 2), Vgf (VGF nerve growth factor inducible) and Wif1 (WNT inhibitory factor 1).

Differential expression of the GDNF receptors Gfra1, Gfra2 and Ret has been reported previously in postnatal rat thalamus, with notable overlap in the reticular nucleus [104], also seen in embryonic mouse [105]. These proteins have well defined roles in enteric nervous system formation but their functions in developing thalamus have not been elucidated. Selective expression of Neuritin-1 and VGF across thalamic

nuclei has also been noted previously and these secreted molecules may play a role in regulating cortical differentiation in areas innervated by these thalamic axons [106]. Igfbp5, Inhba, Tgfb2 and Wif1 all have known roles in differentiation, proliferation and apoptosis in the developing brain but have not been directly implicated in axon guidance or synaptic connectivity. There are no described roles for Bmp3 in nervous system development.
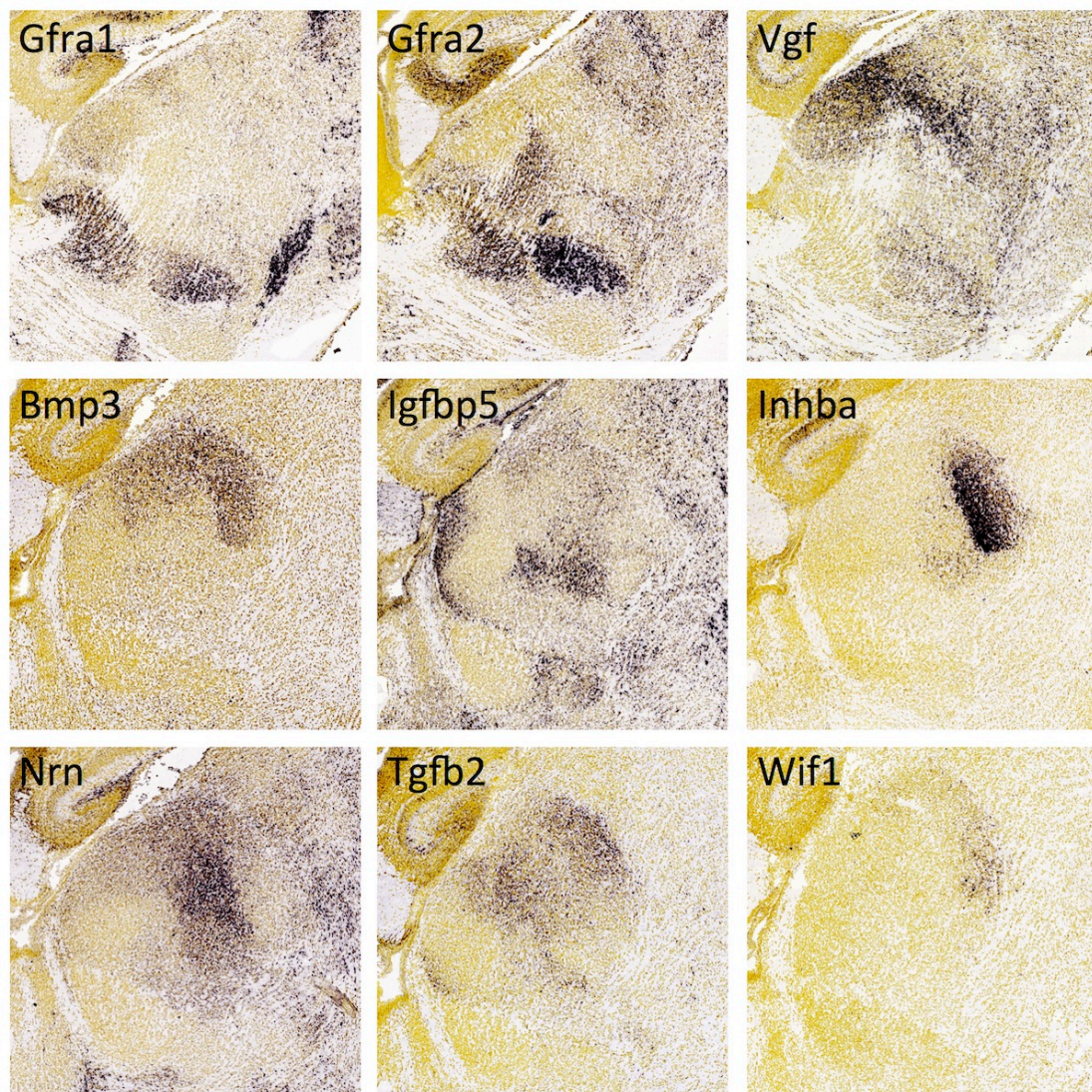
Figure 20. Expression of growth factors and receptors in the thalamus at E18.5. *In situ* hybridization data from sagittal sections of the thalamus at E18.5 obtained from the devABA revealed differential patterns of expression of Gfra1 and Gfra2 (GDNF Family Receptor Alpha 1 and 2), Vgf (VGF nerve growth factor inducible), Bmp3 (Bone morphogenetic protein-3), Igfbp5 (Insulin-like growth factor binding protein 5), Inhba (Inhibin/Activin, beta A subunit), Nrn (Neuritin or Neuritin-1/Nrn1),

Tgfb2 (Transforming growth factor, beta 2) and Wif1 (WNT inhibitory factor 1). See text for details. A representative section is shown for each gene (see Figure S7 for details on how the sections were selected). Gfra1, Gfra2 and Vgf expression is shown laterally, while the other gene expression patterns are shown medially.

Discussion

Our dual screening approach has identified 82 genes encoding candidate connectivity labels in the developing thalamus. Several lines of evidence make these molecules highly plausible candidates to be involved in specifying some aspects of thalamic connectivity: (i) they encode proteins with motifs that are common in known connectivity molecules; (ii) some of them are members of gene families with already known functions in axon guidance or synaptic connectivity; (iii) they cluster with known axon guidance or synaptogenesis molecules based on their temporal pattern of expression in the embryonic thalamus; and (iv) they are selectively or differentially expressed across dorsal thalamic nuclei at stages when connectivity is being established.

The genes we have identified include multiple members of large families where some genes have been previously implicated, such as Ephrins and Eph receptors, cadherins and protocadherins, semaphorins and plexins, and Odz/teneurin genes, where we have expanded the number of candidates potentially involved. Our findings also highlight the selective expression in thalamus of a number of genes or gene families recently implicated in synaptic connectivity in other regions, including several Ig and LRR family members, calsyntenins, cerebellins and Netrin-Gs. In addition, we have found strikingly selective expression of many genes from diverse families not previously implicated in either thalamocortical development or neuronal connectivity.

The systematic nature of the devABA data provides the opportunity to collate and compare these expression patterns across gene families, and to draw some general inferences about the molecular logic specifying thalamic connectivity. The most striking aspect of these patterns is how unique they all are. Across these 82 genes, even when considering only one or two anatomical sections, we did not observe any two patterns that we would say are identical. Moreover, most of the differentiation of expression levels (at E15.5 or E18.5) reflects discrete nuclear divisions, rather than

broad gradients across the whole dorsal thalamus, which, a priori, was just as likely an outcome. This indicates that thalamic nuclei are highly differentiated from each other, with each one displaying a unique repertoire of these molecules. There is, in addition, clear evidence of non-uniform gene expression within nuclei, reflecting the known sub-differentiation of these fields. Thus, while initial guidance of thalamic axons may be defined by general topography [reviewed in 9], later connectivity decisions are likely regulated by combinations of more spatially discrete labels.

There is indeed strong evidence for combinatorial interactions, both within and between the gene families identified. For example, co-expression of Ephrins and Eph-receptors in cis can alter responses to these proteins in trans [107]. Similar modulatory interactions in cis have been observed for Class 6 Semaphorins and Plexin-A proteins, thus generating combinatorial functional diversity [108-110]. Cadherins and protocadherins can also form heteromeric complexes, either within or across these two subfamilies [111,112], which can alter function, as with a cis interaction between Pcdh19 and Cdh2, which generates a novel trans-adhesive complex [113]. Ig superfamily members of the L1 and Cntn families can mediate homophilic adhesion, but also modulate signaling of semaphorin and Ephrin pathways in cis [71,114], while Cntns and Cntnaps also form functional complexes that regulate connectivity [reviewed in 115,116]. Multiple proteins in our dataset also interact with Neuroligin-Neurexin complexes implicated in synaptic development, including cerebellins [90], calsyntenins [39], Igsf9b [25], Mdga1 [78,79] and Gpc3 [96]. This pathway may thus represent a convergence point for combinatorial functions of multiple regulators of synaptogenesis, also including diverse members of the LRR superfamily and LAR protein tyrosine phosphatases not surveyed here [reviewed in 115,116].

Whether the genes we have identified actually encode connectivity labels will, of course, require functional experiments, but the list here is at least likely highly enriched for such molecules. It is by no means comprehensive, however, as both of our strategies had important limitations. The first approach concentrated on genes that are conserved from vertebrates to invertebrates, based on the simple rationale that many of the major families of axon guidance molecules are conserved. Such molecules may be expected to play more important and thus more obvious roles in nervous system development than newly evolved ones, with concomitantly stronger

phenotypes when mutated. In addition, functional analyses of such gene families may be further simplified in invertebrate model systems, which often only have one gene copy, whereas mammals often have multiple genes with possibly redundant functions. Thus, while this strategy identified plausible genes that are highly amenable to functional analyses, it excluded the large fraction of the mammalian proteome that is not obviously conserved in invertebrates.

Our second strategy took advantage of the Allen Brain Atlas database of expression in the developing mouse brain [117]. This database enabled a systematic and quantitative comparison of expression profiles across seven embryonic and postnatal stages, which highlighted several clusters enriched for known or putative connectivity labels. We integrated annotations on protein topology, structural motifs and localization to better characterize the genes in this dataset and enable us to recognise possible surface labels. However, the dataset is limited to 2002 genes, which were selected on the basis of functional relevance to brain development and neurodevelopmental disorders, and which fall into several categories: (i) transcription factors, (ii) known regional or cell-type-specific markers, (iii) neurotransmitters and receptors, (iv) genes associated with patterning or axon guidance signaling pathways and (v) highly studied common drug targets, including GPCRs, ion channels, cell adhesion molecules and genes associated with neurodevelopmental disorders. While this set is thus itself likely enriched for connectivity labels, it is also certain that many others are not represented. For example, extracellular leucine-rich repeat (LRR) proteins are a major class of neuronal connectivity labels [118], with 135 such genes in the mouse genome [23], but only a small number are included in the devABA dataset.

In summary, the extensive set of genes identified here as candidate connectivity labels provides a strong starting point and context for functional analyses of single molecules and of their potential combinatorial interactions, which may coordinately specify the complex connectivity patterns of thalamic nuclei.

Materials and Methods

Mouse *in situ* hybridization

Digoxigenin (DIG)-labelled antisense cRNA probes for *in situ* hybridization were designed to encompass a section of coding sequence and 3'UTR >500bp in length. Briefly this involved TA cloning of PCR products into the TOPO vector (Invitrogen) and simultaneous synthesis and DIG-labelling of RNA transcripts from linearised vector using T7- or Sp6- RNA polymerase. Some cRNA probes were transcribed directly from the PCR products amplified with a sense primer and an antisense primer with T7 promoter sequence at its 5'-end. Primers used were:

Kirrel1_for TGACTATGGCACAGAGCCTA

Kirrel1_rev GAAGGAGAGGAAAGGGCA

Kirrel2_for CAAAAGAACTTGGTCCGGAT

Kirrel2_rev CACTTGGAGAGATGGATTCAC

Kirrel3_for TACGTGCAGTTTGACAAGGC

Kirrel3_rev AGGCTGCAAGGAATACAGAC

Igsf9_for AGACTCACCTCCTGCAAATC

Igsf9_T7Rev GCGGTAATACGACTCACTATAGGGCATTCCTGTTCAGCTCCCA

Igsf9b_for TGACACATTCCACAACGG

Igsf9b_T7Rev GCGGTAATACGACTCACTATAGGGCCCTTATTCCACTTCACCACAG

Sdk1_for CATCATACACTGTGGACCTG

Sdk1_T7rev GCGGTAATACGACTCACTATAGGGAGAAGTTCTGCACCCGTC

Sdk2_for AATGGTGGTTCTTAGTGGTCA

Sdk2_T7rev GCGGTAATACGACTCACTATAGGGCAAACGAAGGATGAGAACC

Odz1_for CATTCAGTGTGAGCTCCAGA

Odz1_T7rev GCGGTAATACGACTCACTATAGGGCTGACGCAAAGGCAGAGAT

Odz1_2_for TGAGGTCCAATATGAGATCC

Odz1_2_T7rev GCGGTAATACGACTCACTATAGGGCAACATCATAATCTCTTTGCCC

Odz2_for AGAGTCAAGCAAGCGAGAA

Odz2_T7rev GCGGTAATACGACTCACTATAGGGCGAGGTGGCAGCAGGTTAT

Odz2_2_for GTGCAATATGAGATGTTCCG

Odz2_2_T7rev GCGGTAATACGACTCACTATAGGGCAGTAAAGTGGACGAGCTTGG

Odz3_for GAAGAGTCAACAGTGGGAAGA

Odz3_T7rev GCGGTAATACGACTCACTATAGGGCTGCTACAGGAGAATCTGCAC

Odz3_2_for CAACCAAATCATTTCCACG

Odz3_2_T7rev GCGGTAATACGACTCACTATAGGGCTGGATTAGTTTGGTGAGCG

Odz4_for TTCAGAAGCAACTCAAGGC

Odz4_T7rev GCGGTAATACGACTCACTATAGGGCCAACACAGTCAGGAATACGG

Odz4_2_for AGAAGGAGCTGAAGGTGG

Odz4_2_T7rev GCGGTAATACGACTCACTATAGGGCCACGATGCTGTTGCTACTC

Neto1_for GGCTCCAGAACTGTGTATATCC

Neto1_T7rev GCGGTAATACGACTCACTATAGGGCGAGTCTTGCCGAAGGAATA

Neto1_2_for AGAAGTCAGTGCAGTGTGG

Neto1_2_T7rev GCGGTAATACGACTCACTATAGGGCGAGCTGCTTCCGTCATAG

Neto2_for CATCAGGAATTGTCTTGGTC

Neto2_T7rev GCGGTAATACGACTCACTATAGGGCCTGCTCTGCCTGACTTAACA

Neto2_2_for GAATCAAGCACATTCCTGC

Neto2_2_T7rev GCGGTAATACGACTCACTATAGGGCGGATCACTCCGACTCCTG

Clstn1_for CACCCTGAACATCGATCC

Clstn1_T7rev GCGGTAATACGACTCACTATAGGGCTCACCCTCTGTCGACAAG

Clstn2_for GCCACTGTCGTCATCATTATC

Clstn2_T7rev GCGGTAATACGACTCACTATAGGGCATGGTGACATCATCTCCAGA

Clstn3_for TCCCAGTCATGTGCTCAGT

Clstn3_T7rev GCGGTAATACGACTCACTATAGGGCAGTAGAGAAGGACACAGCAGC

In situ hybridization was carried out on vibratome-sectioned C57BL6 mouse brains (Jackson Laboratories) as described in [119] with modifications. To obtain brains at E15, timed pregnancy mice were sacrificed by cervical dislocation and embryos were fixed and stored in 4% paraformaldehyde (PFA)-PBS at 4°C until use. For P0, brains were dissected out prior to immersion in 4% PFA-PBS -at 4°C. P9 brains were fixed by perfusion followed by immersion in 4% PFA-PBS at 4°C. Fixed brains were embedded in 3% agarose and 70-μm sections were obtained on a vibratome (VT1000S Leica). Sections were washed twice in PBST (PBS containing 0.1% Tween-20), permeabilised in RIPA buffer (150mM NaCl, 50mM Tris-HCl pH 8.0, 1mM EDTA, 1% Nonidet-P40, 0.5% sodium deoxycholate, 0.1% SDS) and postfixed in PBS containing 4% PFA and 0.2% gluteraldehyde. Hybridisation was performed in a humidified environment overnight at 65°C with 1μg/ml labeled probe in

hybridisation buffer (50% formamide, 5X SSC pH4.5, 1% SDS, 50μg/ml yeast tRNA, 50μg/ml heparin). Posthybridisation washes were completed at 65°C using solution I (50% formamide, 5X SSC pH4.5, 1% SDS) and solution III (50% formamide, 2X SSC pH4.5, 0.1% Tween-20) and at room temperature using TBST (TBS containing 1% Tween-20). Brain sections were incubated for >1hr in blocking buffer (TBST, 10% heat-inactivated sheep serum). Immunodetection was carried out in blocking buffer at 4°C overnight using an alkaline phosphatase-conjugated anti-digoxigenin antibody (Roche) at a 1:2000 dilution. Following antibody incubation extensive TBST washes were performed. Sections were equilibrated in NTMT (100mM Tris-HCl pH 9.5, 100mM NaCl, 50mM $MgCl_2$, 1% Tween-20) prior to colourimetric detection using 2μl/ml NBT/BCIP (Roche) in NTMT. Sections were mounted on Superfrost glass slides (VWR international) and analysed with an Olympus IX51 microscope.

Database pipeline

The database pipeline was used as described in [23]. In summary, the Ensembl database, the International Protein Index and the Mammalian Gene Collection were used to build proteomes for human, mouse, worm and fly. Through automated and manual curation duplicates were removed and only the longest isoforms were kept. This resulted in non-redundant data sets for mouse, human, worm, and fly containing 85991, 74866, 22698, and 16857 sequences, respectively. From an all-against-all Blast search the top 200 hits were used as input for Markov clustering with the MCL program. The output was combined with rich annotation, including predictions of motifs and GPI links, as well as transmembrane and signalling sequences.

Data acquisition and preprocessing

The Allen Brain Atlas (ABA) provides a survey of gene expression in the brain from mid-gestation through aging in different species. Of interest to us is the data on the mouse brain. The detailed process of mouse brain data acquisition is explained in Lein et al. [120]. Briefly, *in situ* hybridization (ISH) data was recorded from coronal or sagittal sections and quantified per unit from a given spatial location for each gene (expression values per voxels). Voxels were assigned regional labels through age-

matched anatomic reference atlases. The developmental dataset on the ABA portal (www.brain-map.org) comprises expression values of 2002 genes obtained from ISH images in the sagittal plane across four embryonic (E11.5, E13.5, E15.5 and E18.5) and three early postnatal ages (P4, P14 and P28) [117]. These time-points are developmentally relevant to regional specification/patterning (E11.5), axon pathfinding (E13.5, E15.5 and E18.5), synaptogenesis (P4 and P14) and cortical plasticity (P28). One animal was used per time-point per gene. Per-region data was quantified from ISH images by combining all voxels with the same regional label. Liscovitch and Chechik [121] extracted data from 36 anatomically delineated regions of the developing brain, which encompass the entire brain. The data is readily available for download at http://chechiklab.biu.ac.il/~lior/cerebellum.html. The data is expressed as expression density. For each brain region R, the expression density is defined as the sum of expressing pixels in R divided by the total number of pixels that intersect R (taken from the Technical White Paper: Informatics Data Processing for the Allen Developing Brain Atlas found at http://developingmouse.brain-map.org/docs/InformaticsDataProcessing.pdf). Each set of values is dependent on specific gene expression. We normalized the data to reflect variations of expression for each gene. A value of 1 indicates average expression levels whereas a value below or above 1 indicates decreased or increased expression levels compared to average, respectively. With this normalised data, we generated the developmental matrix of 1996 genes x 7 time-points as 6 genes had no expression values throughout. These genes are Chrna1, Hoxd9, Lef1, Nrp2, Pbx1 and Rnd2.

The genes of the developmental dataset fall into one of these five categories (taken from the Technical White Paper: Allen Developing Mouse Brain Atlas found at http://developingmouse.brain-map.org/docs/Overview.pdf):

1. *Transcription factors*. Approximately one third of the genes are transcription factors, with extensive coverage of homeobox, basic helix-loop-helix, forkhead, nuclear receptor, high mobility group and POU domain genes.

2. *Neuropeptides, neurotransmitters, and their receptors*. Extensive coverage of genes in dopaminergic, serotonergic, glutamatergic, and gabaergic signaling, as well as neuropeptides and their receptors.

3. *Neuroanatomical marker genes*. Characterizing region- or cell type specific marker

genes over development can provide information about the origins of a brain region or cell type, and may help to identify precursor regions at earlier time points.

4. *Gene ontologies/signaling pathways relevant to brain development*. Gene ontologies include axon guidance, receptor tyrosine kinases and their ligands. Pathways include Wnt signaling and Notch signaling pathways.

5. *Genes of general interest*. This category includes highly studied genes such as common drug targets, ion channels, cell adhesion, genes involved in neurotransmission, G-protein-coupled receptors, or involved in neurodevelopmental diseases, which are expressed in the brain in the adult and/or in development.

We screened protein products for cellular localization with an emphasis on the extracellular compartment. More specifically, we looked for signal peptides, transmembrane domains and glycophosphatidylinositol (GPI)-anchors. Since genes were listed with their Allen gene name in the dataset, we looked for the MGI symbol and Ensembl gene identifiers (through http://www.ncbi.nlm.nih.gov/gene) to subsequently find their protein counterparts. Where multiple isoforms exist, the longest sequence was used for further analyses. We screened Fasta format sequences through PSORTII to find cellular localization of the protein products and the number of predicted helices of transmembrane proteins (http://psort.hgc.jp)[122,123]. PSORTII predicted protein localization in the following compartments: cytoplasm, cytoskeleton, endoplasmic reticulum, Golgi, mitochondria, nucleus, plasma membrane, peroxisomes, vacuoles and vesicles of the secretory system. In addition, we ran the SignalP prediction program to determine whether a protein contains a signal peptide or not - indicative of extracellular localization (http://www.cbs.dtu.dk/services/SignalP/, version 4.1)[124,125]. A GPI modification site prediction, the big-PI predictor was run to identify protein products with potential GPI-anchors (http://mendel.imp.ac.at/sat/gpi/gpi_server.html, version last modified June 17th 2005) [126-129]. When any of the three programs suggested extracellular localization, the protein was deemed found in the extracellular compartment. The TMHMM program was run to assess the presence and number of transmembrane domains (http://www.cbs.dtu.dk/services/TMHMM/, version 2.0)[130,131]. When the number of transmembrane domains varied between PSORTII and TMHMM results, manual curation was performed (unless both programs picked up over two transmembrane domains and the protein was simply considered multi-pass). PFAM

(http://pfam.xfam.org)[132] and SMART (http://smart.embl-heidelberg.de, version 2.12.1 released on August 6th 2012)[133,134], motif recognition programs, allowed screening for important protein domains, which can be amalgamated through InterPro (http://www.ebi.ac.uk/interpro/search/sequence-search, version 49.0 released on November 20th 2014)[135]. For identification of important protein domains we retrieved PFAM, SMART, Interpro and SignalP information through BioMart version 0.7 [136].

Finally, we updated the gene annotations of the dataset from Thompson et al. [117]. Thompson and colleagues had 1385 genes out of the 2002 falling into one or more of these categories: axon guidance pathway, cell adhesion, receptor tyrosine kinases (RTKs) and their ligands, Notch signaling, Wnt signaling, transcription factor activity, basic helix-loop-helix transcription factor, forkhead transcription factor, homeobox transcription factor, nuclear receptor, POU domain genes, neurotransmitter pathway, ion channel, and G-protein coupled receptors (GPCRs). We screened genes for the following gene ontology (GO) term annotations (http://www.geneontology.org)[137] through MartView:

1) Axon Guidance GO:0007411;

2) Synapse GO:0045202;

3) Patterning: - Regionalization GO:0003002;

     - Pattern Specification Process GO:0007389;

     - Neural Tube Patterning GO:0021532;

     - Rostrocaudal Neural Tube Patterning GO:0021903;

4) Transcription Factor Activity:

   - Protein Binding Transcription Factor Activity GO:0000988;

   - Nucleic Acid Binding Transcription Factor Activity GO:0001071;

   - Sequence-Specific DNA Binding Transcription Factor Activity GO:0003700.

Since many genes had more than one annotation, we grouped them as follows:
- Group 1: axon guidance pathway and cell adhesion;
- Group 2: synapse;
- Group 3: receptor tyrosine kinases and their ligands, and patterning;

- Group 4: neurotransmission pathway (GPCRs, ion channels, gap junctions);

- Group 5: chromatin and transcription factor activity

- Group 6: other (cytoskeleton, extracellular matrix, myelin, metabolic enzymes and signal transduction) and unannotated.

We only had 206 unannotated genes. Each gene belonged to only one group. When a gene fell into more than one group, we prioritised the groups as presented.

Clustering analyses

We performed *k*-means clustering with the Euclidean distance method on the developmental dataset using the clustering module integrated in ArrayPipe (http://www.pathogenomics.ca/arraypipe - version 1.7)[138]. Each clustering analysis was iterated a thousand times.

Enrichment analyses

We performed chi square statistics to determine whether our developmental dataset was normally distributed in regards to cellular localisation and groups across the 10 clusters obtained during *k*-means clustering analyses or enriched in particular clusters. Statistical significance was determined at p-value < 0.05. Where multiple tests were performed on the same dataset, the Bonferroni correction was taken into account.

Acknowledgments

References

1. Sherman SM (2007) The thalamus is more than just a relay. Curr Opin Neurobiol 17: 417-422.

2. Smith Y, Wichmann T (2015) The cortico-pallidal projection: an additional route for cortical regulation of the basal ganglia circuitry. Mov Disord 30: 293-295.

3. McHaffie JG, Stanford TR, Stein BE, Coizet V, Redgrave P (2005) Subcortical loops through the basal ganglia. Trends Neurosci 28: 401-407.

4. Bostan AC, Dum RP, Strick PL (2013) Cerebellar networks with the cerebral cortex and basal ganglia. Trends Cogn Sci 17: 241-254.

5. Lam YW, Sherman SM (2011) Functional organization of the thalamic input to the thalamic reticular nucleus. J Neurosci 31: 6791-6799.

6. Clasca F, Rubio-Garrido P, Jabaudon D (2012) Unveiling the diversity of thalamocortical neuron subtypes. Eur J Neurosci 35: 1524-1532.

7. Rubio-Garrido P, Perez-de-Manzo F, Porrero C, Galazo MJ, Clasca F (2009) Thalamic input to distal apical dendrites in neocortical layer 1 is massive and highly convergent. Cereb Cortex 19: 2380-2395.

8. Garel S, Lopez-Bendito G (2014) Inputs from the thalamocortical system on axon pathfinding mechanisms. Curr Opin Neurobiol 27: 143-150.

9. Molnar Z, Garel S, Lopez-Bendito G, Maness P, Price DJ (2012) Mechanisms controlling the guidance of thalamocortical axons through the embryonic forebrain. Eur J Neurosci 35: 1573-1585.

10. Osterhout JA, Josten N, Yamada J, Pan F, Wu SW, et al. (2011) Cadherin-6 mediates axon-target matching in a non-image-forming visual circuit. Neuron 71: 632-639.

11. Su J, Klemm MA, Josephson AM, Fox MA (2013) Contributions of VLDLR and LRP8 in the establishment of retinogeniculate projections. Neural Dev 8: 11.

12. Leamey CA, Merlin S, Lattouf P, Sawatari A, Zhou X, et al. (2007) Ten_m3 regulates eye-specific patterning in the mammalian visual pathway and is required for binocular vision. PLoS Biol 5: e241.

13. Tran H, Sawatari A, Leamey CA (2015) The glycoprotein Ten-m3 mediates topography and patterning of thalamostriatal projections from the parafascicular nucleus in mice. Eur J Neurosci 41: 55-68.

14. Nakagawa Y, Shimogori T (2012) Diversity of thalamic progenitor cells and postmitotic neurons. Eur J Neurosci 35: 1554-1562.

15. Suzuki-Hirano A, Ogawa M, Kataoka A, Yoshida AC, Itoh D, et al. (2011) Dynamic spatiotemporal gene expression in embryonic mouse thalamus. J Comp Neurol 519: 528-543.

16. Price DJ, Clegg J, Duocastella XO, Willshaw D, Pratt T (2012) The importance of combinatorial gene expression in early Mammalian thalamic patterning and thalamocortical axonal guidance. Front Neurosci 6: 37.

17. Jones EG, Rubenstein JL (2004) Expression of regulatory genes during differentiation of thalamic nuclei in mouse and monkey. J Comp Neurol 477: 55-80.

18. Nakagawa Y, O'Leary DD (2001) Combinatorial expression patterns of LIM-homeodomain and other regulatory genes parcellate developing thalamus. J Neurosci 21: 2711-2725.

19. Lehigh KM, Leonard CE, Baranoski J, Donoghue MJ (2013) Parcellation of the thalamus into distinct nuclei reflects EphA expression and function. Gene Expr Patterns 13: 454-463.

20. Yuge K, Kataoka A, Yoshida AC, Itoh D, Aggarwal M, et al. (2011) Region-specific gene expression in early postnatal mouse thalamus. J Comp Neurol 519: 544-561.

21. Marcos-Mondejar P, Peregrin S, Li JY, Carlsson L, Tole S, et al. (2012) The lhx2 transcription factor controls thalamocortical axonal guidance by specific regulation of robo1 and robo2 receptors. J Neurosci 32: 4372-4385.

22. Chatterjee M, Li JY (2012) Patterning and compartment formation in the diencephalon. Front Neurosci 6: 66.

23. Dolan J, Walshe K, Alsbury S, Hokamp K, O'Keeffe S, et al. (2007) The extracellular leucine-rich repeat superfamily; a comparative survey and analysis of evolutionary relationships and expression patterns. BMC Genomics 8: 320.

24. Hansen M, Walmod PS (2013) IGSF9 family proteins. Neurochem Res 38: 1236-1251.

25. Woo J, Kwon SK, Nam J, Choi S, Takahashi H, et al. (2013) The adhesion protein IgSF9b is coupled to neuroligin 2 via S-SCAM to promote inhibitory synapse development. J Cell Biol 201: 929-944.

26. Mishra A, Traut MH, Becker L, Klopstock T, Stein V, et al. (2014) Genetic evidence for the adhesion protein IgSF9/Dasm1 to regulate inhibitory synapse development independent of its intracellular domain. J Neurosci 34: 4187-4199.

27. Fischbach KF, Linneweber GA, Andlauer TF, Hertenstein A, Bonengel B, et al. (2009) The irre cell recognition module (IRM) proteins. J Neurogenet 23: 48-67.

28. Shen K, Fetter RD, Bargmann CI (2004) Synaptic specificity is generated by the synaptic guidepost protein SYG-2 and its receptor, SYG-1. Cell 116: 869-881.

29. Prince JE, Brignall AC, Cutforth T, Shen K, Cloutier JF (2013) Kirrel3 is required for the coalescence of vomeronasal sensory neuron axons into glomeruli and for male-male aggression. Development 140: 2398-2408.

30. Martin EA, Muralidhar S, Wang Z, Cervantes DC, Basu R, et al. (2015) The intellectual disability gene Kirrel3 regulates target-specific mossy fiber synapse development in the hippocampus. Elife 4: e09395.

31. Nguyen DN, Liu Y, Litsky ML, Reinke R (1997) The sidekick gene, a member of the immunoglobulin superfamily, is required for pattern formation in the Drosophila eye. Development 124: 3303-3312.

32. Krishnaswamy A, Yamagata M, Duan X, Hong YK, Sanes JR (2015) Sidekick 2 directs formation of a retinal circuit that detects differential motion. Nature 524: 466-470.

33. Yamagata M, Sanes JR (2008) Dscam and Sidekick proteins direct lamina-specific synaptic connections in vertebrate retina. Nature 451: 465-469.

34. Mosca TJ (2015) On the Teneurin track: a new synaptic organization molecule emerges. Front Cell Neurosci 9: 204.

35. Leamey CA, Sawatari A (2014) The teneurins: new players in the generation of visual topography. Semin Cell Dev Biol 35: 173-179.

36. Copits BA, Swanson GT (2012) Dancing partners at the synapse: auxiliary subunits that shape kainate receptor function. Nat Rev Neurosci 13: 675-686.

37. Kim YJ, Bao H, Bonanno L, Zhang B, Serpe M (2012) Drosophila Neto is essential for clustering glutamate receptors at the neuromuscular junction. Genes Dev 26: 974-987.

38. Hintsch G, Zurlinden A, Meskenaite V, Steuble M, Fink-Widmer K, et al. (2002) The calsyntenins--a family of postsynaptic membrane proteins with distinct neuronal expression patterns. Mol Cell Neurosci 21: 393-409.

39. Um JW, Pramanik G, Ko JS, Song MY, Lee D, et al. (2014) Calsyntenins function as synaptogenic adhesion molecules in concert with neurexins. Cell Rep 6: 1096-1109.

40. Pfeiffenberger C, Yamada J, Feldheim DA (2006) Ephrin-As and patterned retinal activity act together in the development of topographic maps in the primary visual system. J Neurosci 26: 12873-12884.

41. Dufour A, Seibt J, Passante L, Depaepe V, Ciossek T, et al. (2003) Area specificity and topography of thalamocortical projections are controlled by ephrin/Eph genes. Neuron 39: 453-465.

42. Uziel D, Garcez P, Lent R, Peuckert C, Niehage R, et al. (2006) Connecting thalamus and cortex: the role of ephrins. Anat Rec A Discov Mol Cell Evol Biol 288: 135-142.

43. Tai AX, Kromer LF (2014) Corticofugal projections from medial primary somatosensory cortex avoid EphA7-expressing neurons in striatum and thalamus. Neuroscience 274: 409-418.

44. Torii M, Levitt P (2005) Dissociation of corticothalamic and thalamocortical axon targeting by an EphA7-mediated mechanism. Neuron 48: 563-575.

45. Gerstmann K, Pensold D, Symmank J, Khundadze M, Hubner CA, et al. (2015) Thalamic afferents influence cortical progenitors via ephrin A5-EphA4 interactions. Development 142: 140-150.

46. Robichaux MA, Chenaux G, Ho HY, Soskis MJ, Dravis C, et al. (2014) EphB receptor forward signaling regulates area-specific reciprocal thalamic and cortical axon pathfinding. Proc Natl Acad Sci U S A 111: 2188-2193.

47. Yun ME, Johnson RR, Antic A, Donoghue MJ (2003) EphA family gene expression in the developing mouse neocortex: regional patterns reveal intrinsic programs and extrinsic influence. J Comp Neurol 456: 203-216.

48. Suzuki SC, Inoue T, Kimura Y, Tanaka T, Takeichi M (1997) Neuronal circuits are subdivided by differential expression of type-II classic cadherins in postnatal mouse brains. Mol Cell Neurosci 9: 433-447.

49. Poskanzer K, Needleman LA, Bozdagi O, Huntley GW (2003) N-cadherin regulates ingrowth and laminar targeting of thalamocortical axons. J Neurosci 23: 2294-2305.

50. Gil OD, Needleman L, Huntley GW (2002) Developmental patterns of cadherin expression and localization in relation to compartmentalized thalamocortical terminations in rat barrel cortex. J Comp Neurol 453: 372-388.

51. Kim SY, Chung HS, Sun W, Kim H (2007) Spatiotemporal expression pattern of non-clustered protocadherin family members in the developing rat brain. Neuroscience 147: 996-1021.

52. Uemura M, Nakao S, Suzuki ST, Takeichi M, Hirano S (2007) OL-Protocadherin is essential for growth of striatal axons and thalamocortical projections. Nat Neurosci 10: 1151-1159.

53. Carcea I, Patil SB, Robison AJ, Mesias R, Huntsman MM, et al. (2014) Maturation of cortical circuits requires Semaphorin 7A. Proc Natl Acad Sci U S A 111: 13978-13983.

54. Fukunishi A, Maruyama T, Zhao H, Tiwari M, Kang S, et al. (2011) The action of Semaphorin7A on thalamocortical axon branching. J Neurochem 118: 1008-1015.

55. Leighton PA, Mitchell KJ, Goodrich LV, Lu X, Pinson K, et al. (2001) Defining brain wiring patterns and mechanisms through gene trapping in mice. Nature 410: 174-179.

56. Little GE, Lopez-Bendito G, Runker AE, Garcia N, Pinon MC, et al. (2009) Specificity and plasticity of thalamocortical connections in Sema6A mutant mice. PLoS Biol 7: e98.

57. Demyanenko GP, Riday TT, Tran TS, Dalal J, Darnell EP, et al. (2011) NrCAM deletion causes topographic mistargeting of thalamocortical axons to the visual cortex and disrupts visual acuity. J Neurosci 31: 1545-1558.

58. Wright AG, Demyanenko GP, Powell A, Schachner M, Enriquez-Barreto L, et al. (2007) Close homolog of L1 and neuropilin 1 mediate guidance of thalamocortical axons at the ventral telencephalon. J Neurosci 27: 13667-13679.

59. Guijarro P, Wang Y, Ying Y, Yao Y, Jieyi X, et al. (2013) In vivo knockdown of cKit impairs neuronal migration and axonal extension in the cerebral cortex. Dev Neurobiol 73: 871-887.

60. Widenfalk J, Nosrat C, Tomac A, Westphal H, Hoffer B, et al. (1997) Neurturin and glial cell line-derived neurotrophic factor receptor-beta (GDNFR-beta), novel proteins related to GDNF and GDNFR-alpha with specific cellular patterns of expression suggesting roles in the developing and adult nervous system and in peripheral organs. J Neurosci 17: 8506-8519.

61. Lush ME, Ma L, Parada LF (2005) TrkB signaling regulates the developmental maturation of the somatosensory cortex. Int J Dev Neurosci 23: 523-536.

62. Cheng J, Wu K, Armanini M, O'Rourke N, Dowbenko D, et al. (1997) A novel protein-tyrosine phosphatase related to the homotypically adhering kappa and mu receptors. J Biol Chem 272: 7264-7277.

63. Taniguchi Y, London R, Schinkmann K, Jiang S, Avraham H (1999) The receptor protein tyrosine phosphatase, PTP-RO, is upregulated during megakaryocyte differentiation and Is associated with the c-Kit receptor. Blood 94: 539-549.

64. Johnson KG, Van Vactor D (2003) Receptor protein tyrosine phosphatases in nervous system development. Physiol Rev 83: 1-24.

65. Weiner JA, Koo SJ, Nicolas S, Fraboulet S, Pfaff SL, et al. (2004) Axon fasciculation defects and retinal dysplasias in mice lacking the immunoglobulin superfamily adhesion molecule BEN/ALCAM/SC1. Mol Cell Neurosci 27: 59-69.

66. Buhusi M, Demyanenko GP, Jannie KM, Dalal J, Darnell EP, et al. (2009) ALCAM regulates mediolateral retinotopic mapping in the superior colliculus. J Neurosci 29: 15630-15641.

67. Fogel AI, Akins MR, Krupp AJ, Stagi M, Stein V, et al. (2007) SynCAMs organize synapses through heterophilic adhesion. J Neurosci 27: 12516-12530.

68. Frei JA, Stoeckli ET (2014) SynCAMs extend their functions beyond the synapse. Eur J Neurosci 39: 1752-1760.

69. Toth AB, Terauchi A, Zhang LY, Johnson-Venkatesh EM, Larsen DJ, et al. (2013) Synapse maturation by activity-dependent ectodomain shedding of SIRPalpha. Nat Neurosci 16: 1417-1425.

70. Hsieh CP, Chang WT, Lee YC, Huang AM (2015) Deficits in cerebellar granule cell development and social interactions in CD47 knockout mice. Dev Neurobiol 75: 463-484.

71. Shimoda Y, Watanabe K (2009) Contactins: emerging key roles in the development and function of the nervous system. Cell Adh Migr 3: 64-70.

72. Sakurai K, Toyoshima M, Ueda H, Matsubara K, Takeda Y, et al. (2009) Contribution of the neural cell recognition molecule NB-3 to synapse formation between parallel fibers and Purkinje cells in mouse. Dev Neurobiol 69: 811-824.

73. Sakurai K, Toyoshima M, Takeda Y, Shimoda Y, Watanabe K (2010) Synaptic formation in subsets of glutamatergic terminals in the mouse hippocampal formation is affected by a deficiency in the neural cell recognition molecule NB-3. Neurosci Lett 473: 102-106.

74. Lee S, Takeda Y, Kawano H, Hosoya H, Nomoto M, et al. (2000) Expression and regulation of a gene encoding neural recognition molecule NB-3 of the contactin/F3 subgroup in mouse brain. Gene 245: 253-266.

75. Takeda Y, Akasaka K, Lee S, Kobayashi S, Kawano H, et al. (2003) Impaired motor coordination in mice lacking neural recognition molecule NB-3 of the contactin/F3 subgroup. J Neurobiol 56: 252-265.

76. Takeuchi A, Hamasaki T, Litwack ED, O'Leary DD (2007) Novel IgCAM, MDGA1, expressed in unique cortical area- and layer-specific patterns and transiently by distinct forebrain populations of Cajal-Retzius neurons. Cereb Cortex 17: 1531-1541.

77. Ishikawa T, Gotoh N, Murayama C, Abe T, Iwashita M, et al. (2011) IgSF molecule MDGA1 is involved in radial migration and positioning of a subset of cortical upper-layer neurons. Dev Dyn 240: 96-107.

78. Lee K, Kim Y, Lee SJ, Qiang Y, Lee D, et al. (2013) MDGAs interact selectively with neuroligin-2 but not other neuroligins to regulate inhibitory synapse development. Proc Natl Acad Sci U S A 110: 336-341.

79. Pettem KL, Yokomaku D, Takahashi H, Ge Y, Craig AM (2013) Interaction between autism-linked MDGAs and neuroligins suppresses inhibitory synapse development. J Cell Biol 200: 321-336.

80. de Wit J, Hong W, Luo L, Ghosh A (2011) Role of leucine-rich repeat proteins in the development and function of neural circuits. Annu Rev Cell Dev Biol 27: 697-729.

81. Kegel L, Aunin E, Meijer D, Bermingham JR (2013) LGI proteins in the nervous system. ASN Neuro 5: 167-181.

82. Fukata Y, Lovero KL, Iwanaga T, Watanabe A, Yokoi N, et al. (2010) Disruption of LGI1-linked synaptic complex causes abnormal synaptic transmission and epilepsy. Proc Natl Acad Sci U S A 107: 3799-3804.

83. Fukata Y, Adesnik H, Iwanaga T, Bredt DS, Nicoll RA, et al. (2006) Epilepsy-related ligand/receptor complex LGI1 and ADAM22 regulate synaptic transmission. Science 313: 1792-1795.

84. Seppala EH, Jokinen TS, Fukata M, Fukata Y, Webster MT, et al. (2011) LGI2 truncation causes a remitting focal epilepsy in dogs. PLoS Genet 7: e1002194.

85. Andreae LC, Lumsden A, Gilthorpe JD (2009) Chick Lrrn2, a novel downstream effector of Hoxb1 and Shh, functions in the selective targeting of rhombomere 4 motor neurons. Neural Dev 4: 27.

86. Tossell K, Andreae LC, Cudmore C, Lang E, Muthukrishnan U, et al. (2011) Lrrn1 is required for formation of the midbrain-hindbrain boundary and organiser through regulation of affinity differences between midbrain and hindbrain cells in chick. Dev Biol 352: 341-352.

87. Aruga J (2003) Slitrk6 expression profile in the mouse embryo and its relationship to that of Nlrr3. Gene Expr Patterns 3: 727-733.

88. Dickendesher TL, Baldwin KT, Mironova YA, Koriyama Y, Raiker SJ, et al. (2012) NgR1 and NgR3 are receptors for chondroitin sulfate proteoglycans. Nat Neurosci 15: 703-712.

89. Wills ZP, Mandel-Brehm C, Mardinly AR, McCord AE, Giger RJ, et al. (2012) The nogo receptor family restricts synapse number in the developing hippocampus. Neuron 73: 466-481.

90. Uemura T, Lee SJ, Yasumura M, Takeuchi T, Yoshida T, et al. (2010) Trans-synaptic interaction of GluRdelta2 and Neurexin through Cbln1 mediates synapse formation in the cerebellum. Cell 141: 1068-1079.

91. Wei P, Pattarini R, Rong Y, Guo H, Bansal PK, et al. (2012) The Cbln family of proteins interact with multiple signaling pathways. J Neurochem 121: 717-729.

92. Miura E, Iijima T, Yuzaki M, Watanabe M (2006) Distinct expression of Cbln family mRNAs in developing and adult mouse brains. Eur J Neurosci 24: 750-760.

93. Matsukawa H, Akiyoshi-Nishimura S, Zhang Q, Lujan R, Yamaguchi K, et al. (2014) Netrin-G/NGL complexes encode functional synaptic diversification. J Neurosci 34: 15779-15792.

94. Nakashiba T, Nishimura S, Ikeda T, Itohara S (2002) Complementary expression and neurite outgrowth activity of netrin-G subfamily members. Mech Dev 111: 47-60.

95. Karayannis T, Au E, Patel JC, Kruglikov I, Markx S, et al. (2014) Cntnap4 differentially contributes to GABAergic and dopaminergic synaptic transmission. Nature 511: 236-240.

96. de Wit J, O'Sullivan ML, Savas JN, Condomitti G, Caccese MC, et al. (2013) Unbiased discovery of glypican as a receptor for LRRTM4 in regulating excitatory synapse development. Neuron 79: 696-711.

97. Budnik V, Salinas PC (2011) Wnt signaling during synaptic development and plasticity. Curr Opin Neurobiol 21: 151-159.

98. Sahores M, Gibb A, Salinas PC (2010) Frizzled-5, a receptor for the synaptic organizer Wnt7a, regulates activity-mediated synaptogenesis. Development 137: 2215-2225.

99. Liu C, Wang Y, Smallwood PM, Nathans J (2008) An essential role for Frizzled5 in neuronal survival in the parafascicular nucleus of the thalamus. J Neurosci 28: 5641-5653.

100. Wilson PM, Fryer RH, Fang Y, Hatten ME (2010) Astn2, a novel member of the astrotactin gene family, regulates the trafficking of ASTN1 during glial-guided neuronal migration. J Neurosci 30: 8529-8540.

101. Garcia-Gallastegui P, Ibarretxe G, Garcia-Ramirez JJ, Baladron V, Aurrekoetxea M, et al. (2014) DLK1 regulates branching morphogenesis and parasympathetic innervation of salivary glands through inhibition of NOTCH signalling. Biol Cell 106: 237-253.

102. Fukazawa N, Yokoyama S, Eiraku M, Kengaku M, Maeda N (2008) Receptor type protein tyrosine phosphatase zeta-pleiotrophin signaling controls endocytic trafficking of DNER that regulates neuritogenesis. Mol Cell Biol 28: 4494-4506.

103. Dessaud E, Salaun D, Gayet O, Chabbert M, deLapeyriere O (2006) Identification of lynx2, a novel member of the ly-6/neurotoxin superfamily, expressed in neuronal subpopulations during mouse development. Mol Cell Neurosci 31: 232-242.

104. Burazin TC, Gundlach AL (1999) Localization of GDNF/neurturin receptor (c-ret, GFRalpha-1 and alpha-2) mRNAs in postnatal rat brain: differential regional and temporal expression in hippocampus, cortex and cerebellum. Brain Res Mol Brain Res 73: 151-171.

105. Golden JP, DeMaro JA, Osborne PA, Milbrandt J, Johnson EM, Jr. (1999) Expression of neurturin, GDNF, and GDNF family-receptor mRNA in the developing and mature mouse. Exp Neurol 158: 504-528.

106. Sato H, Fukutani Y, Yamamoto Y, Tatara E, Takemoto M, et al. (2012) Thalamus-derived molecules promote survival and dendritic growth of developing cortical neurons. J Neurosci 32: 15388-15402.

107. Kao TJ, Kania A (2011) Ephrin-mediated cis-attenuation of Eph receptor signaling is essential for spinal motor axon guidance. Neuron 71: 76-91.

108. Sun LO, Brady CM, Cahill H, Al-Khindi T, Sakuta H, et al. (2015) Functional assembly of accessory optic system circuitry critical for compensatory eye movements. Neuron 86: 971-984.

109. Haklai-Topper L, Mlechkovich G, Savariego D, Gokhman I, Yaron A (2010) Cis interaction between Semaphorin6A and Plexin-A4 modulates the repulsive response to Sema6A. EMBO J 29: 2635-2645.

110. Perez-Branguli F, Zagar Y, Shanley DK, Graef IA, Chedotal A, et al. (2016) Reverse Signaling by Semaphorin-6A Regulates Cellular Aggregation and Neuronal Morphology. PLoS One 11: e0158686.

111. Sotomayor M, Gaudet R, Corey DP (2014) Sorting out a promiscuous superfamily: towards cadherin connectomics. Trends Cell Biol 24: 524-536.

112. Basu R, Taylor MR, Williams ME (2015) The classic cadherins in synaptic specificity. Cell Adh Migr 9: 193-201.

113. Emond MR, Biswas S, Blevins CJ, Jontes JD (2011) A complex of Protocadherin-19 and N-cadherin mediates a novel mechanism of cell adhesion. J Cell Biol 195: 1115-1121.

114. Schmid RS, Maness PF (2008) L1 and NCAM adhesion molecules as signaling coreceptors in neuronal migration and process outgrowth. Curr Opin Neurobiol 18: 245-250.

115. Yang X, Hou D, Jiang W, Zhang C (2014) Intercellular protein-protein interactions at synapses. Protein Cell 5: 420-444.

116. de Wit J, Ghosh A (2016) Specification of synaptic connectivity by cell surface interactions. Nat Rev Neurosci 17: 22-35.

117. Thompson CL, Ng L, Menon V, Martinez S, Lee CK, et al. (2014) A high-resolution spatiotemporal atlas of gene expression of the developing mouse brain. Neuron 83: 309-323.

118. de Wit J, Ghosh A (2014) Control of neural circuit formation by leucine-rich repeat proteins. Trends Neurosci 37: 539-550.

119. Riddle RD, Johnson RL, Laufer E, Tabin C (1993) Sonic hedgehog mediates the polarizing activity of the ZPA. Cell 75: 1401-1416.

120. Lein ES, Hawrylycz MJ, Ao N, Ayres M, Bensinger A, et al. (2007) Genome-wide atlas of gene expression in the adult mouse brain. Nature 445: 168-176.

121. Liscovitch N, Chechik G (2013) Specialization of gene expression during mouse brain development. PLoS Comput Biol 9: e1003185.

122. Horton P, Nakai K (1997) Better prediction of protein cellular localization sites with the k nearest neighbors classifier. Proc Int Conf Intell Syst Mol Biol 5: 147-152.

123. Nakai K, Horton P (1999) PSORT: a program for detecting sorting signals in proteins and predicting their subcellular localization. Trends Biochem Sci 24: 34-36.

124. Nielsen H, Engelbrecht J, Brunak S, von Heijne G (1997) Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. Protein Eng 10: 1-6.

125. Petersen TN, Brunak S, von Heijne G, Nielsen H (2011) SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat Methods 8: 785-786.

126. Eisenhaber B, Bork P, Eisenhaber F (1998) Sequence properties of GPI-anchored proteins near the omega-site: constraints for the polypeptide binding site of the putative transamidase. Protein Eng 11: 1155-1161.

127. Eisenhaber B, Bork P, Eisenhaber F (1999) Prediction of potential GPI-modification sites in proprotein sequences. J Mol Biol 292: 741-758.

128. Eisenhaber B, Bork P, Yuan Y, Loffler G, Eisenhaber F (2000) Automated annotation of GPI anchor sites: case study C. elegans. Trends Biochem Sci 25: 340-341.

129. Sunyaev SR, Eisenhaber F, Rodchenkov IV, Eisenhaber B, Tumanyan VG, et al. (1999) PSIC: profile extraction from sequence alignments with position-specific counts of independent observations. Protein Eng 12: 387-394.

130. Krogh A, Larsson B, von Heijne G, Sonnhammer EL (2001) Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol 305: 567-580.

131. Sonnhammer EL, von Heijne G, Krogh A (1998) A hidden Markov model for predicting transmembrane helices in protein sequences. Proc Int Conf Intell Syst Mol Biol 6: 175-182.

132. Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, et al. (2014) Pfam: the protein families database. Nucleic Acids Res 42: D222-230.

133. Letunic I, Doerks T, Bork P (2014) SMART: recent updates, new developments and status in 2015. Nucleic Acids Res.

134. Schultz J, Milpetz F, Bork P, Ponting CP (1998) SMART, a simple modular architecture research tool: identification of signaling domains. Proc Natl Acad Sci U S A 95: 5857-5864.

135. Mulder NJ, Apweiler R, Attwood TK, Bairoch A, Bateman A, et al. (2002) InterPro: an integrated documentation resource for protein families, domains and functional sites. Brief Bioinform 3: 225-235.

136. Kasprzyk A (2011) BioMart: driving a paradigm change in biological data management. Database (Oxford) 2011: bar049.

137. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, et al. (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet 25: 25-29.

138. Hokamp K, Roche FM, Acab M, Rousseau ME, Kuo B, et al. (2004) ArrayPipe: a flexible processing pipeline for microarray data. Nucleic Acids Res 32: W457-459.

Supplemental Tables

Table S1. Conserved candidate connectivity labels. Genes are organized in clusters defined by TRIBE-MCL, as described [23]. Genes from Drosophila melanogaster are in blue, from Caenorhabditis elegans in green and from mouse in black. Signal peptides were predicted using SignalP, transmembrane domains by a consensus between TMHMM and HMMTOP and protein motifs by SMART and Pfam, also as described [23].

Table S2. Developmental dataset. Absolute values of expression densities for 2002 genes from the devABA database per gene per time-point are shown.

Table S3. Developmental dataset and *k*-means clustering. Normalised expression densities per gene per time-point (columns B-H). Thalamic expression densities were obtained for 1996 genes. Heatmap's 3 colour scale of gene expression data: 0.2, red; 1, white; blue, 5. Normalised gene expression was clustered by similar patterns of expression using *k*-means Euclidean distance 6-18, with 1000 iterations (columns I-U). Clusters are sorted at $k = 10$, followed by $k = 11$, followed by $k = 12$ and so on until $k = 18$. Three colour scale of cluster number: red < yellow < green, for better visualisation.

Table S4. Developmental dataset protein characteisation. Each gene was matched to its MGI symbol and Ensembl gene identifier. From these, protein sequences were obtained - where multiple isoforms were found, the longest sequence was used for further analyses. Fasta format sequences were screened through PSORT II for cellular localization (PSORT II prediction and probability: gol, Golgi; cyt, cytoplasmic; csk, cytoskeletal; end, endoplasmic reticulum; ext, extracellular, including cell wall; mit, mitochondrial; nuc, nuclear; per, peroxisomal; pla, plasma membrane; ves, vesicles of secretory system). PSORT II uses two-fold *k-nearest* neighbor (k-NN) algorithm with $k1 = 9$ and $k2 = 23$ [122]. Screens for number of predicted transmembrane domains (TMs) with PSORT II were also performed. SignalP detected signal peptides (sigp) on proteins. Big-PI predictor tested different potential GPI-modification sites and returned the GPI best score, GPI profile score, GPI profile independent score and GPI quality [126-129]. GPI quality of P or S indicated a potential GPI-modification site (P: predicted; S: second predicted site, more than one predicted site). TMHMM screened for the presence (Transmembrane Domain) and number of transmembrane

domains (PredHel) [130,131]. Columns AE, AF and AG show SMART ID, PFAM ID and InterPro short description, respectively. Genes were screened for additional GO terms (see methods).

Table S5. Developmental dataset protein localisation and functional annotations. Normalised expression densities per gene per time-point (Columns B-H). Heatmap's 3 colour scale of gene expression data: 0.2, red; 1, white; 5, blue. Data is sorted by cluster number at $k = 10$ followed by gene name (Column I). Each gene was matched to its MGI symbol (Column J). Protein localisation was assessed after combining the data from the different screens performed as either in the Golgi (gol), cytoplasmic (cyt), cytoskeletal (csk), in the endoplasmic reticulum (end), extracellular including cell wall (ext), mitochondrial (mit), nuclear (nuc), peroxisomal (per), on the plasma membrane (pla) or in vesicles of secretory system (ves; Column K). Moreover, when the protein was deemed in the extracellular compartment, it was found to be either GPI-achored (anchored), secreted, single-pass transmembrane (single-pass) or multi-pass transmembrane (multi-pass; Column L). Functional assessment was grouped in six different categories: 1) axon guidance pathway and cell adhesion; 2) synapse; 3) receptor tyrosine kinases and their ligands, and patterning; 4) neurotransmission pathway (GPCRs, ion channels, gap junctions); 5) chromatin and transcription factor activity; and 6) other (cytoskeleton, extracellular matrix, myelin, metabolic enzymes and signal transduction) and unannotated (Column M). Three colour scale of cluster number and group number: red < yellow < green, for better visualisation (Columns I and M). See text for details.

Table S6. Known and potential connectivity genes. Genes are listed by cluster and grouped into those with known roles in thalamocortical connectivity, known functions in axon guidance of synaptogenesis more generally or no such known functions (Other).

Supplemental Figures

Figure S1. *In situ* hybridization patterns for immunoglobulin superfamily genes across entire brain at P0. Two coronal sections are shown for Kirrel3, Igsf9b and Sdk2, one rostral and one more caudal.

Figure S2. *In situ* **hybridization patterns for Odz family genes across entire brain at P0.** Two coronal sections are shown for Odz1, Odz2, Odz3 and Odz4, one rostral and one more caudal. In addition to restricted expression in dorsal thalamus, there is also graded expression of Odz genes across cortex (cx) and striatum (str) in differing patterns.
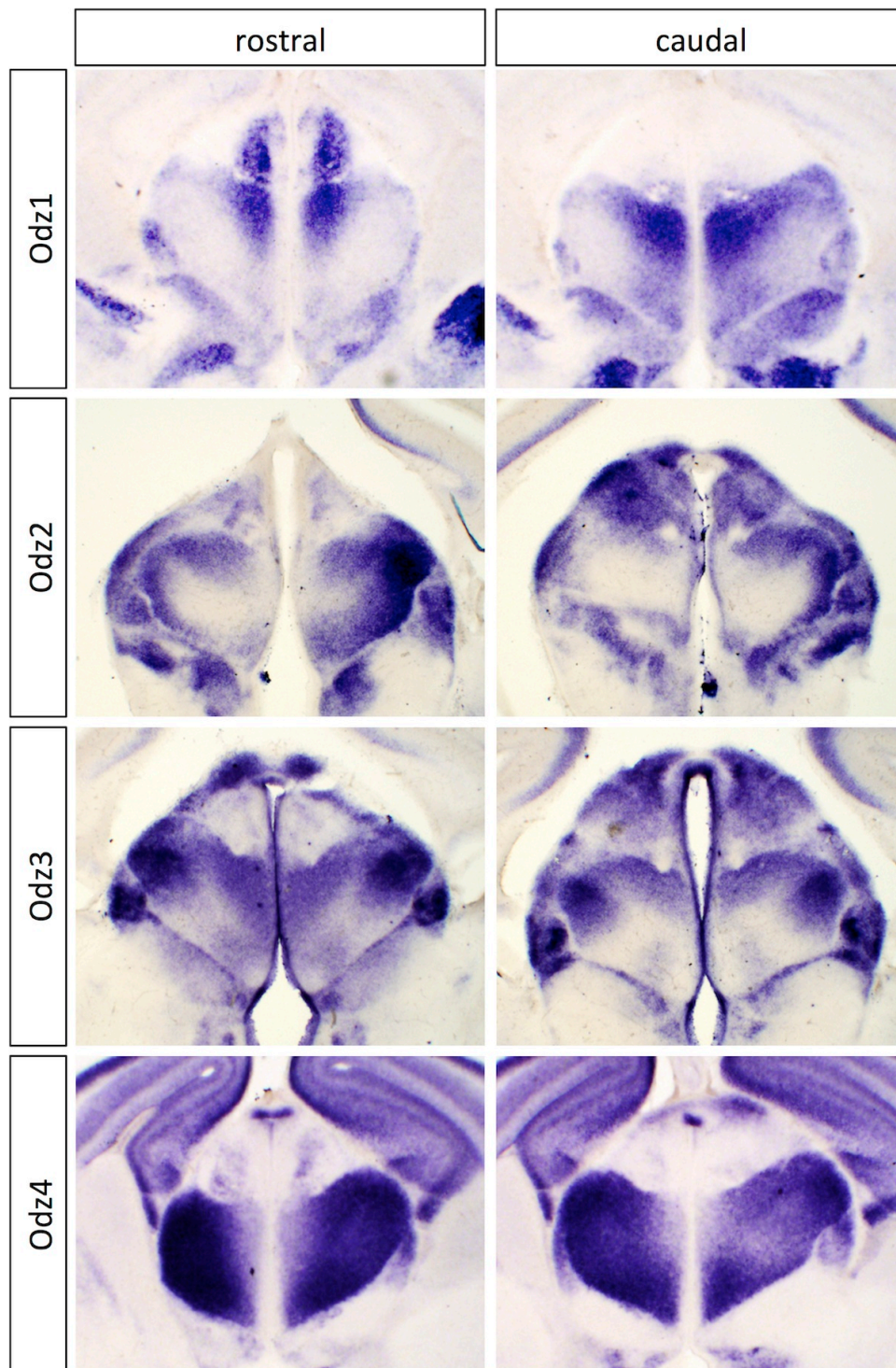
Figure S3. *In situ* hybridization patterns for Odz family genes in dorsal thalamus at E15.5. Two coronal sections are shown for Odz1, Odz2, Odz3 and Odz4, one rostral and one more caudal. Differential expression across the dorsal thalamus is already evident at this stage. The corresponding entire brain sections are shown in Figure S4.

Figure S4. *In situ* hybridization patterns for Odz family genes across entire brain at E15.5. Two coronal sections are shown for Odz1, Odz2, Odz3 and Odz4, one rostral and one more caudal.

Figure S5. *In situ* **hybridization patterns for Neto family genes across entire brain at P0.** Three coronal sections are shown for Neto1 and Neto2, one rostral, one at an intermediate level (mid) and one more caudal.
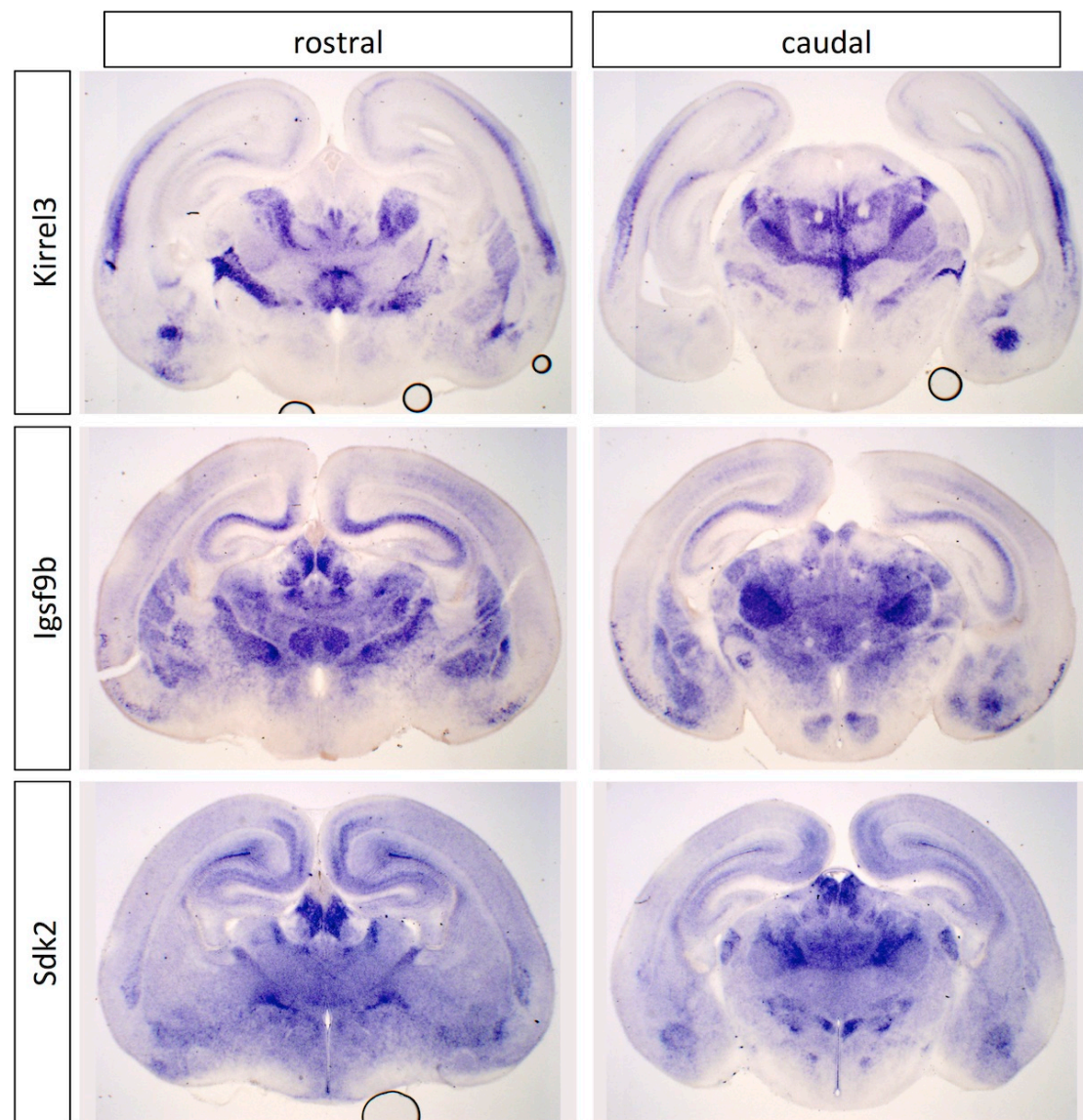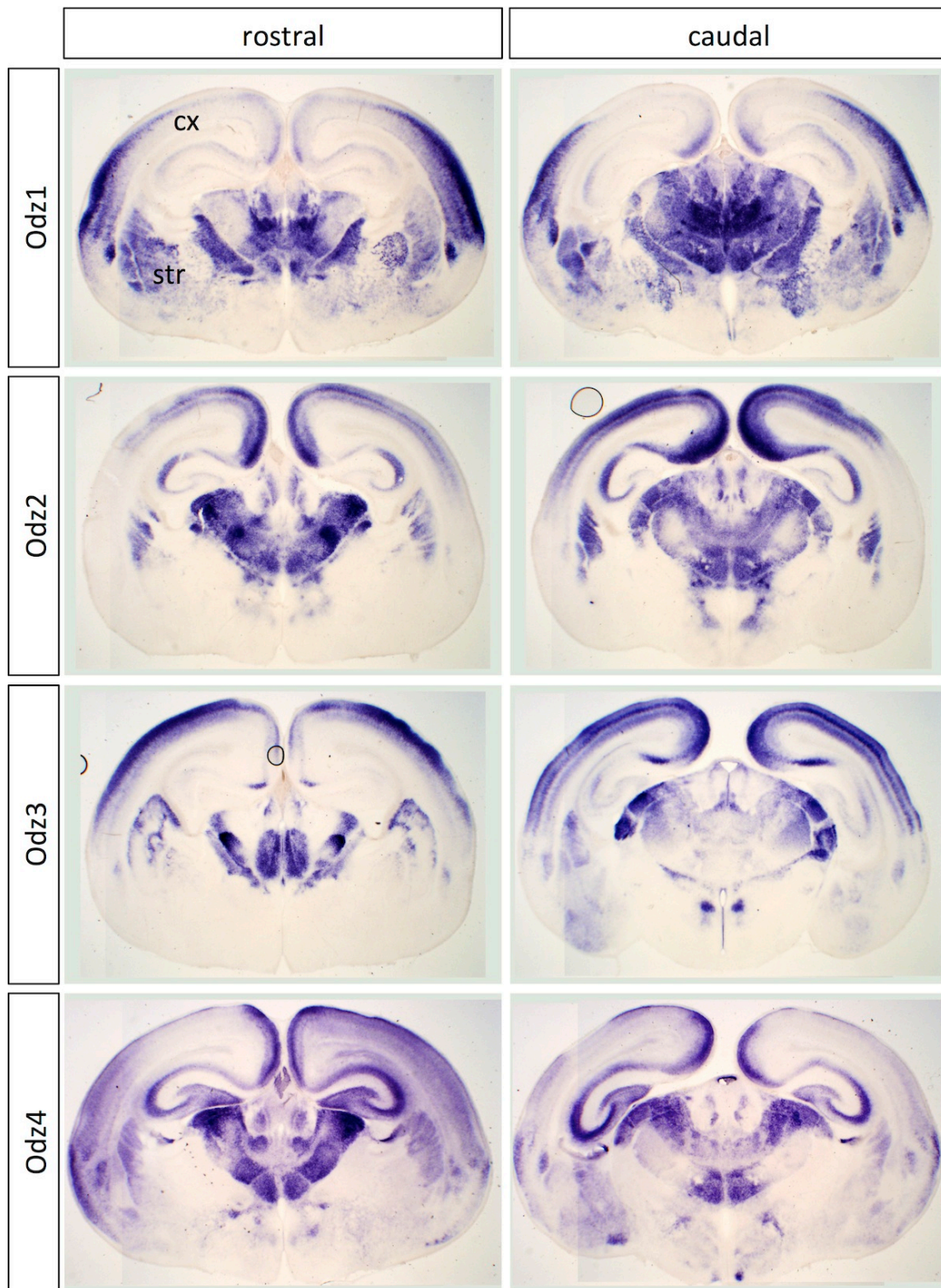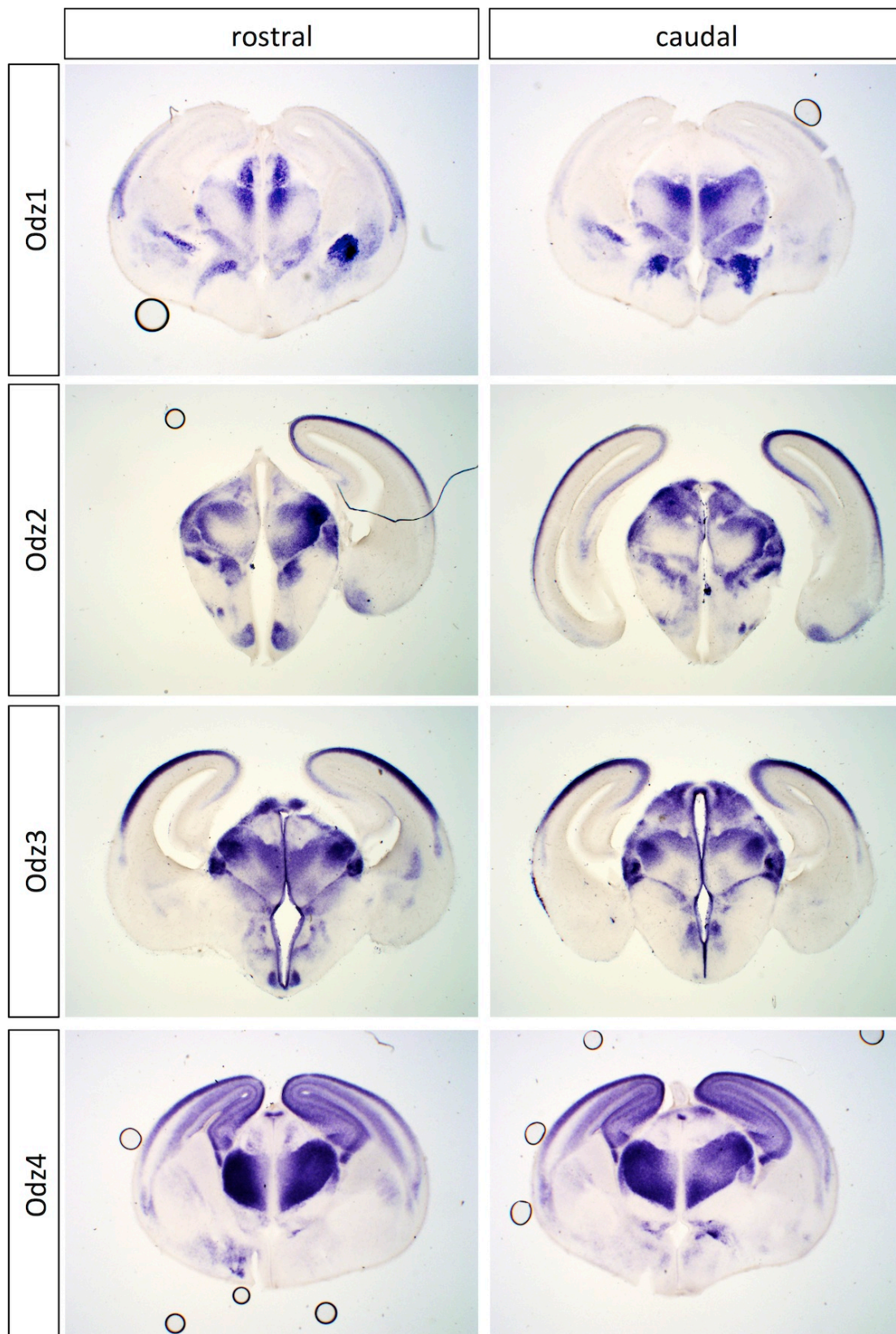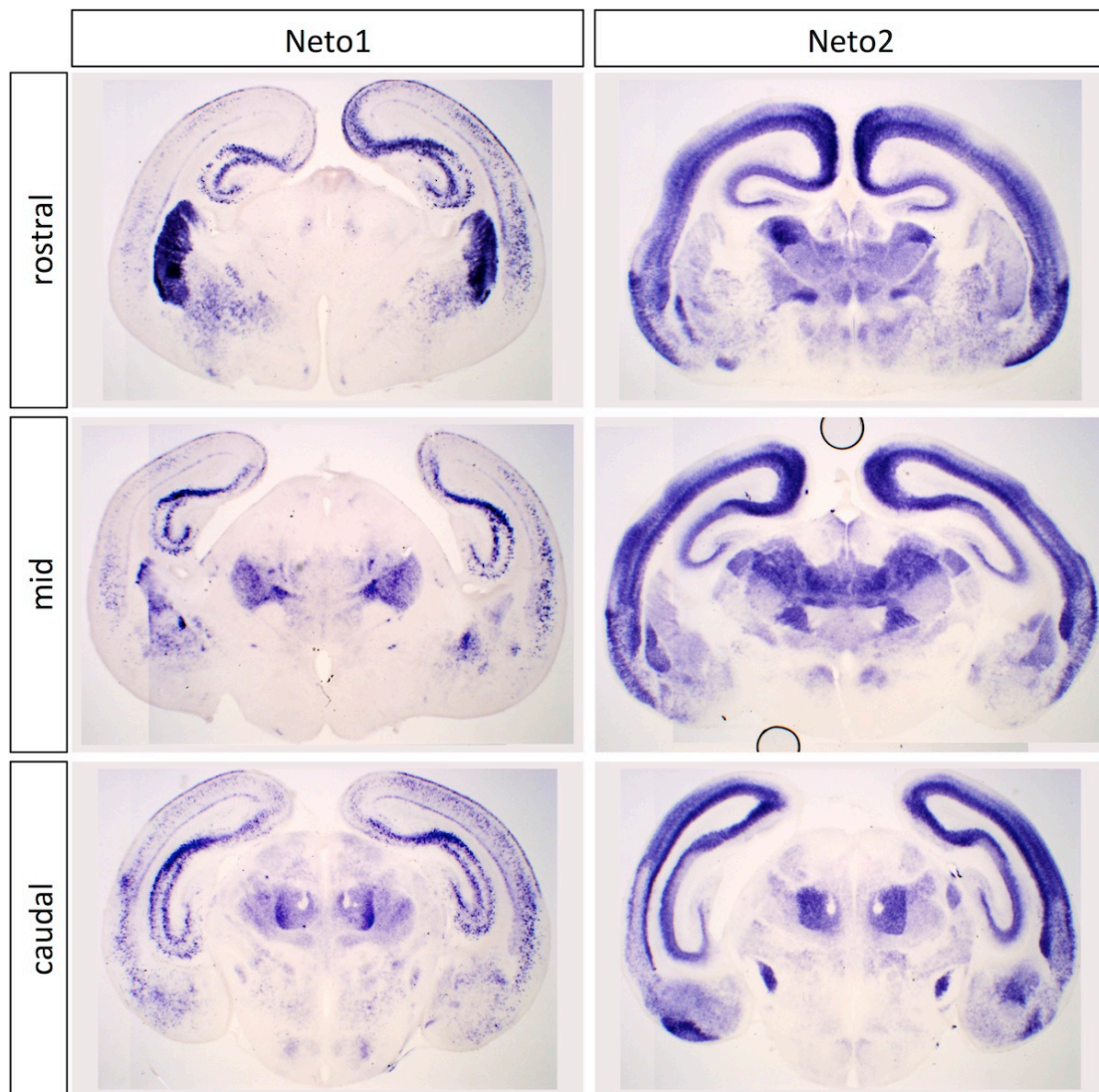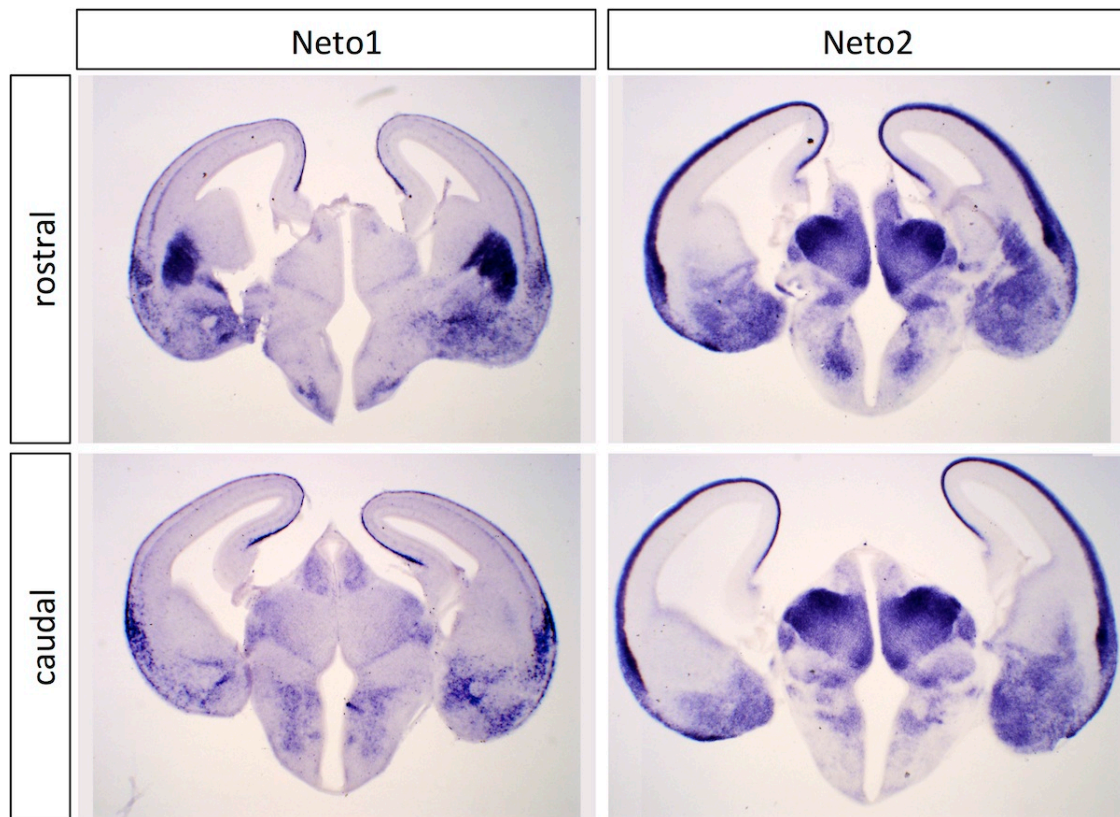
Figure S6. *In situ* hybridization patterns for Neto family genes across entire brain at E15.5. Two coronal sections are shown for Neto1 and Neto2, one rostral and one more caudal. Differential expression across the dorsal thalamus is already evident at this stage.

k-mean 6	E11.5	E13.5	E15.5	E18.5	P04	P14	P28	# genes
cluster 5	3.325	0.647	0.361	0.620	0.598	0.807	0.577	326
cluster 0	1.046	1.680	1.046	0.763	0.806	1.024	0.630	300
cluster 2	0.627	0.427	0.544	2.194	1.235	1.212	0.771	307
cluster 1	0.387	0.283	0.356	0.804	2.777	1.474	0.917	392
cluster 3	0.453	0.275	0.252	0.525	0.629	3.684	1.157	414
cluster 4	0.342	0.245	0.248	0.427	0.752	1.984	2.966	257

k-mean 7	E11.5	E13.5	E15.5	E18.5	P04	P14	P28	# genes
cluster 0	3.333	0.644	0.362	0.633	0.598	0.787	0.575	255
cluster 5	1.068	1.722	1.059	0.809	0.777	0.960	0.590	284
cluster 4	0.637	0.415	0.517	2.548	1.315	0.981	0.584	206
cluster 6	0.425	0.298	0.348	0.786	3.371	1.067	0.678	213
cluster 1	0.431	0.343	0.449	0.877	1.497	2.077	1.306	456
cluster 3	0.442	0.256	0.215	0.455	0.499	3.901	1.199	341
cluster 2	0.372	0.239	0.197	0.393	0.590	1.865	3.290	241

k-mean 8	E11.5	E13.5	E15.5	E18.5	P04	P14	P28	# genes
cluster 5	4.087	0.739	0.318	0.505	0.430	0.478	0.374	133
cluster 1	2.077	0.653	0.484	0.791	0.786	1.411	0.817	273
cluster 0	0.865	1.965	1.170	0.820	0.817	0.798	0.539	208
cluster 3	0.568	0.411	0.533	2.575	1.303	0.989	0.599	196
cluster 7	0.401	0.287	0.345	0.786	3.399	1.063	0.688	206
cluster 6	0.320	0.333	0.445	0.880	1.561	2.095	1.347	420
cluster 4	0.397	0.249	0.216	0.450	0.500	3.933	1.211	329
cluster 2	0.328	0.238	0.198	0.391	0.581	1.874	3.332	231

k-mean 9	E11.5	E13.5	E15.5	E18.5	P04	P14	P28	# genes
cluster 1	4.104	0.748	0.319	0.485	0.415	0.480	0.378	131
cluster 7	2.068	0.652	0.499	0.825	0.801	1.393	0.780	276
cluster 6	0.874	1.990	1.164	0.812	0.791	0.811	0.531	205
cluster 5	0.572	0.405	0.501	2.685	1.302	0.958	0.551	173
cluster 0	0.400	0.279	0.341	0.783	3.445	1.059	0.664	197
cluster 4	0.331	0.345	0.484	0.966	1.618	1.980	1.253	407
cluster 2	0.434	0.260	0.224	0.501	0.502	4.200	0.832	221
cluster 3	0.279	0.213	0.183	0.332	0.655	2.940	2.364	256
cluster 8	0.448	0.298	0.242	0.497	0.571	1.167	3.685	130

k-mean 10	**E11.5**	**E13.5**	**E15.5**	**E18.5**	**P04**	**P14**	**P28**	**# genes**
cluster 7	4.087	0.733	0.318	0.503	0.430	0.478	0.374	131
cluster 6	2.067	0.665	0.465	0.791	0.785	1.423	0.820	268
cluster 5	0.932	2.706	0.674	0.686	0.560	0.946	0.447	106
cluster 2	0.769	0.966	1.548	1.047	1.196	0.769	0.703	163
cluster 4	0.553	0.378	0.426	2.779	1.286	1.003	0.550	157
cluster 0	0.317	0.318	0.421	0.965	1.628	2.069	1.256	384
cluster 8	0.403	0.284	0.303	0.791	3.467	1.070	0.662	192
cluster 9	0.432	0.249	0.226	0.482	0.491	4.237	0.839	214
cluster 3	0.271	0.213	0.190	0.329	0.650	2.949	2.374	253
cluster 1	0.450	0.294	0.225	0.494	0.566	1.175	3.699	128

k-mean 11	E11.5	E13.5	E15.5	E18.5	P04	P14	P28	# genes
cluster 4	4.142	0.757	0.323	0.466	0.405	0.473	0.361	124
cluster 0	2.204	0.643	0.459	0.836	0.846	1.181	0.817	232
cluster 6	0.948	2.756	0.672	0.693	0.600	0.847	0.428	100
cluster 1	0.760	0.982	1.594	1.048	1.172	0.772	0.671	157
cluster 5	0.526	0.383	0.442	2.766	1.291	1.012	0.555	158
cluster 8	0.319	0.298	0.410	0.917	1.870	1.795	1.364	332
cluster 10	0.395	0.257	0.284	0.756	3.635	1.080	0.556	162
cluster 9	0.689	0.550	0.438	0.885	0.718	2.775	0.933	225
cluster 2	0.349	0.187	0.179	0.355	0.444	4.547	0.896	162
cluster 3	0.252	0.155	0.150	0.295	0.684	2.923	2.496	226
cluster 7	0.447	0.302	0.222	0.485	0.520	1.154	3.773	118

k-mean 12	E11.5	E13.5	E15.5	E18.5	P04	P14	P28	# genes
cluster 0	4.127	0.752	0.322	0.465	0.410	0.473	0.380	126
cluster 6	2.251	0.660	0.472	0.850	0.851	1.250	0.663	215
cluster 8	0.911	2.734	0.685	0.688	0.600	0.888	0.421	101
cluster 11	0.777	0.956	1.646	1.069	1.246	0.737	0.568	142
cluster 7	0.526	0.383	0.442	2.766	1.291	1.012	0.555	158
cluster 1	0.412	0.268	0.271	0.739	3.738	0.996	0.536	142
cluster 2	0.283	0.262	0.384	0.910	2.027	1.861	1.251	299
cluster 10	0.629	0.471	0.433	0.894	0.764	2.843	0.927	211
cluster 5	0.354	0.189	0.178	0.350	0.442	4.544	0.904	163
cluster 9	0.209	0.142	0.139	0.280	0.694	2.937	2.549	210
cluster 4	0.837	0.657	0.500	0.844	0.843	1.215	2.081	144
cluster 3	0.380	0.224	0.171	0.319	0.470	1.216	4.133	85

k-mean 13	E11.5	E13.5	E15.5	E18.5	P04	P14	P28	# genes
cluster 6	4.163	0.766	0.322	0.465	0.387	0.469	0.352	121
cluster 9	2.315	0.629	0.443	0.814	0.846	1.178	0.754	210
cluster 11	0.900	2.756	0.674	0.669	0.608	0.888	0.430	98
cluster 5	0.911	1.005	1.770	1.041	1.118	0.673	0.484	117
cluster 12	0.574	0.425	0.431	3.075	0.977	0.989	0.515	107
cluster 8	0.466	0.309	0.480	1.611	2.324	1.094	0.668	171
cluster 1	0.362	0.260	0.251	0.473	3.989	1.042	0.588	104
cluster 3	0.214	0.204	0.291	0.667	1.973	2.218	1.413	239
cluster 0	0.797	0.591	0.446	0.964	0.681	2.731	0.763	179
cluster 4	0.357	0.186	0.186	0.359	0.453	4.558	0.855	158
cluster 10	0.257	0.164	0.165	0.306	0.558	3.026	2.449	217
cluster 2	0.594	0.655	0.633	0.949	1.053	1.330	1.782	177
cluster 7	0.429	0.235	0.169	0.378	0.477	1.213	4.000	98

k-mean 14	E11.5	E13.5	E15.5	E18.5	P04	P14	P28	# genes
cluster 4	4.223	0.743	0.306	0.453	0.360	0.475	0.369	114
cluster 7	2.297	0.726	0.559	1.015	1.067	0.715	0.593	159
cluster 2	2.016	0.652	0.394	0.547	0.515	2.075	0.846	129
cluster 6	0.876	2.792	0.665	0.658	0.620	0.881	0.430	94
cluster 3	0.743	1.034	1.907	0.966	1.110	0.732	0.508	99
cluster 10	0.562	0.421	0.429	3.080	0.969	1.012	0.519	106
cluster 13	0.423	0.309	0.508	1.665	2.288	1.107	0.664	158
cluster 11	0.362	0.255	0.247	0.512	4.018	0.999	0.569	102
cluster 5	0.235	0.218	0.298	0.680	2.078	2.092	1.380	221
cluster 12	0.479	0.464	0.425	0.995	0.819	2.865	0.911	177
cluster 1	0.363	0.187	0.187	0.355	0.447	4.559	0.861	158
cluster 9	0.223	0.156	0.150	0.275	0.603	3.004	2.514	207
cluster 8	0.616	0.651	0.607	0.968	1.056	1.295	1.808	176
cluster 0	0.434	0.234	0.172	0.364	0.478	1.204	4.012	96

k-mean 15	E11.5	E13.5	E15.5	E18.5	P04	P14	P28	# genes
cluster 8	4.231	0.742	0.304	0.451	0.355	0.475	0.370	113
cluster 5	2.319	0.727	0.552	0.995	1.034	0.734	0.603	163
cluster 1	1.964	0.657	0.394	0.517	0.522	2.180	0.831	132
cluster 6	0.882	2.785	0.661	0.657	0.617	0.902	0.430	95
cluster 10	0.784	1.061	1.926	0.969	0.991	0.771	0.502	93
cluster 0	0.562	0.425	0.439	3.140	0.999	0.919	0.508	98
cluster 3	0.455	0.344	0.597	1.620	2.337	0.965	0.630	157
cluster 14	0.367	0.254	0.223	0.469	4.039	1.020	0.588	98
cluster 4	0.240	0.226	0.307	0.731	2.094	2.077	1.302	224
cluster 2	0.447	0.420	0.436	1.395	0.866	2.628	0.819	137
cluster 13	0.365	0.199	0.206	0.373	0.402	4.710	0.707	126
cluster 7	0.332	0.228	0.212	0.352	0.621	3.475	1.714	162
cluster 12	0.190	0.155	0.139	0.261	0.731	2.613	2.863	148
cluster 9	0.633	0.676	0.605	0.948	1.055	1.269	1.814	168
cluster 11	0.475	0.228	0.178	0.390	0.420	1.080	4.108	82

k-mean 16	E11.5	E13.5	E15.5	E18.5	P04	P14	P28	# genes
cluster 7	4.270	0.759	0.304	0.414	0.345	0.476	0.355	108
cluster 6	1.967	0.641	0.408	0.569	0.517	2.144	0.803	127
cluster 4	1.838	1.496	0.590	0.990	0.823	0.769	0.501	154
cluster 10	0.893	2.823	0.671	0.663	0.578	0.874	0.426	91
cluster 5	0.864	1.085	1.979	0.979	0.914	0.680	0.490	85
cluster 1	0.546	0.410	0.428	3.081	1.036	0.971	0.514	108
cluster 3	0.528	0.461	0.642	1.475	2.500	0.763	0.595	129
cluster 11	0.394	0.210	0.179	0.431	4.484	0.795	0.462	59
cluster 2	0.307	0.193	0.237	0.767	2.706	1.919	0.863	145
cluster 14	0.404	0.608	0.675	1.285	1.236	1.677	1.120	191
cluster 13	0.476	0.370	0.346	0.780	0.782	3.216	0.982	182
cluster 0	0.321	0.186	0.160	0.288	0.387	4.771	0.846	122
cluster 12	0.233	0.180	0.286	0.547	1.678	2.126	1.916	159
cluster 15	0.206	0.157	0.129	0.270	0.475	3.086	2.610	166
cluster 9	0.967	0.593	0.453	0.831	0.775	0.930	2.391	89
cluster 8	0.366	0.224	0.167	0.309	0.451	1.248	4.184	81

k-mean 17	E11.5	E13.5	E15.5	E18.5	P04	P14	P28	# genes
cluster 9	4.215	0.751	0.300	0.445	0.351	0.491	0.373	115
cluster 15	2.337	0.734	0.568	1.004	1.015	0.725	0.570	156
cluster 10	1.929	0.633	0.407	0.569	0.526	2.158	0.807	127
cluster 13	0.887	2.816	0.671	0.657	0.602	0.871	0.424	92
cluster 7	0.795	1.081	2.035	0.945	0.961	0.681	0.493	81
cluster 3	0.542	0.414	0.433	3.134	0.966	0.987	0.514	100
cluster 14	0.531	0.385	0.574	1.725	2.349	0.794	0.622	127
cluster 4	0.383	0.298	0.227	0.478	4.283	0.798	0.489	72
cluster 16	0.323	0.217	0.279	0.756	2.657	1.873	0.840	144
cluster 0	0.411	0.622	0.731	1.194	1.203	1.662	1.176	178
cluster 2	0.238	0.189	0.277	0.548	1.734	2.113	1.863	152
cluster 6	0.460	0.405	0.339	1.118	0.830	3.105	0.681	119
cluster 1	0.369	0.191	0.169	0.287	0.381	4.802	0.762	114
cluster 8	0.298	0.208	0.203	0.336	0.600	3.457	1.843	104
cluster 5	0.200	0.178	0.135	0.265	0.516	2.583	3.065	119
cluster 11	0.958	0.602	0.438	0.843	0.788	0.914	2.406	85
cluster 12	0.410	0.221	0.183	0.327	0.438	0.989	4.357	61

k-mean 18	E11.5	E13.5	E15.5	E18.5	P04	P14	P28	# genes
cluster 7	4.276	0.764	0.306	0.404	0.342	0.477	0.353	107
cluster 11	2.495	0.596	0.539	1.018	1.055	0.686	0.573	131
cluster 5	2.029	0.579	0.405	0.571	0.532	2.108	0.836	120
cluster 15	1.223	1.932	0.676	0.880	0.713	1.016	0.528	112
cluster 3	0.722	3.722	0.769	0.436	0.446	0.517	0.335	34
cluster 1	0.752	0.941	2.303	0.956	0.940	0.644	0.444	60
cluster 16	0.539	0.393	0.424	3.069	1.050	0.994	0.523	110
cluster 14	0.565	0.448	0.633	1.466	2.476	0.768	0.603	133
cluster 4	0.376	0.226	0.178	0.437	4.334	0.882	0.524	70
cluster 10	0.288	0.193	0.243	0.808	2.628	1.966	0.858	132
cluster 17	0.391	0.615	0.761	1.210	1.221	1.630	1.172	173
cluster 9	0.236	0.182	0.287	0.575	1.707	2.114	1.865	155
cluster 13	0.524	0.442	0.358	1.122	0.784	3.043	0.681	126
cluster 6	0.356	0.194	0.172	0.290	0.394	4.811	0.744	112
cluster 0	0.315	0.207	0.201	0.337	0.611	3.487	1.793	154
cluster 12	0.191	0.173	0.134	0.260	0.512	2.627	3.045	122
cluster 2	0.934	0.557	0.427	0.834	0.783	0.941	2.455	85
cluster 8	0.415	0.221	0.183	0.314	0.431	0.983	4.375	60

Figure S7. Summaries of expression profiles across *k*-means values. Normalised expression densities were averaged per cluster to see the trends of expression for the results of all clustering analyses from *k* = 6 to 18 (*k* indicated at top left corner of each table). Clusters were organised chronologically with early peaks of expression at the top and later peaks of expression at the bottom. Heatmap's 3 colour scale of gene expression data: 0.2, red; 1, white; 5, blue. *k* = 10 was used for further analyses; the corresponding tableframe is in bold.

Figure S8. Image extraction of thalamic gene expression from the devABA. *In situ* hybridization data from a sagittal section of the thalamus at E18.5 obtained from the devABA (A). Corresponding section from the anatomic reference atlas (B). Higher magnifications of the thalamus from squared regions in A and B, respectively (C and D). p1, prosomere 1 (pretectum and pretectal tegmentum); p1A, alar plate of prosomere 1; p1B, basal plate of prosomere 1; p2, prosomere 2 (thalamus and thalamic tegmentum); p2A, alar plate of prosomere 2; p2B, basal plate of prosomere 2; p3, prosomere 3 (prethalamus and prethalamic tegmentum); p3A, alar plate of prosomere 3; p3B, basal plate of prosomere 3. Note that voxels were assigned regional labels for thalamus and thalamic tegmentum separately.
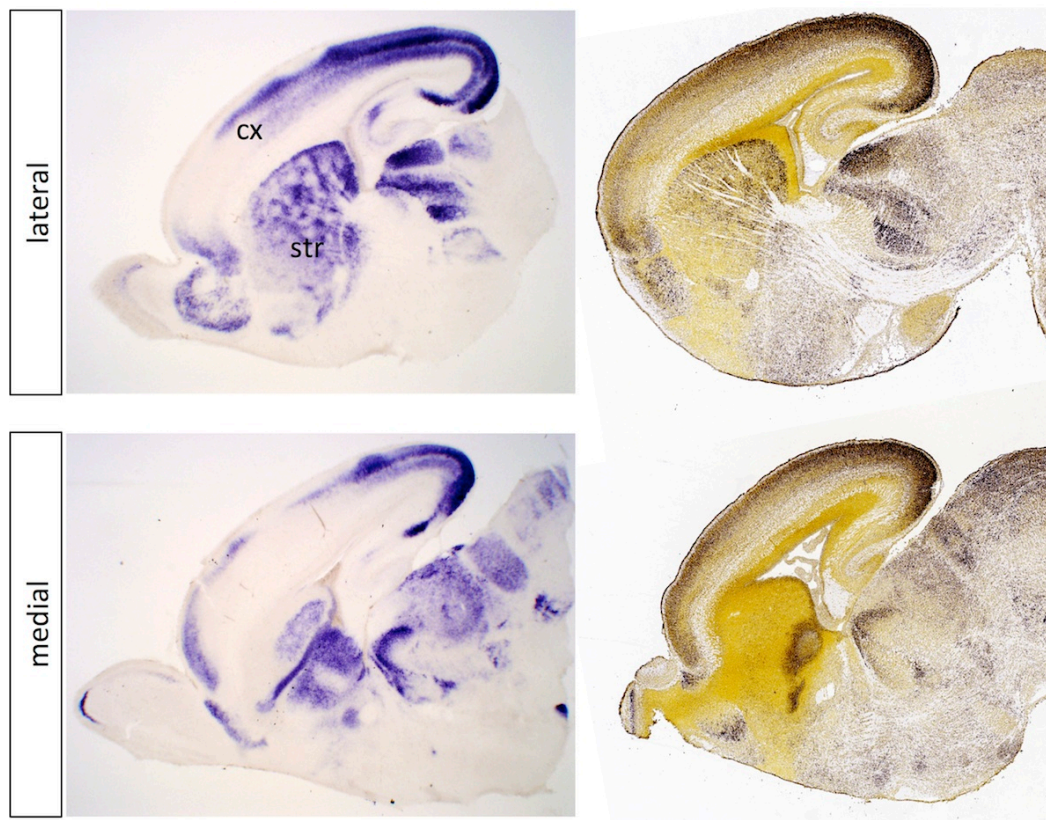
Figure S9. Odz3 *in situ* hybridization compared to devABA data. Two sagittal sections are shown, one lateral and one more medial. Our *in situ* hybridizations are on sections from P0 nenonates, while the devABA sections are from E18.5 embryos. Despite this difference of about a day, there is strikingly good correspondence in expression patterns across the dorsal thalamus, striatum (str), cortex (cx) and other brain regions.