Title: Single cell RNA-Seq analysis identifies molecular mechanisms controlling hypothalamic patterning and differentiation.

Authors: Dong Won Kim[1], Parris Whitney Washington[1*], Zoe Qianyi Wang[1*], Sonia Lin[1], Changyu Sun[1], Lizhi Jiang[1], and Seth Blackshaw[1-6].

Affiliations: [1]Solomon H. Snyder Department of Neuroscience, [2]Department of Ophthalmology, [3]Department of Neurology, [4]Center for Human Systems Biology, [5]Institute for Cell Engineering, [6]Kavli Neuroscience Discovery Institute, Johns Hopkins University School of Medicine, Baltimore, MD, 21205, USA.
*These authors contributed equally.
Address all correspondence to sblack@jhmi.edu (S.B.)

Abstract:

The hypothalamus is a central regulator of many innate behaviors that are essential for survival, but the molecular mechanisms controlling hypothalamic patterning and cell fate specification remain poorly understood. To identify genes that control hypothalamic development, we have used single-cell RNA sequencing (scRNA-Seq) to profile mouse hypothalamic gene expression across 12 developmental time points between embryonic day 10 and postnatal day 45. This identified genes that delineated clear developmental trajectories for all major hypothalamic cell types, and readily distinguished major regional subdivisions of the developing hypothalamus. We show that this approach can rapidly and comprehensively characterize mutants that have altered hypothalamic patterning, and in doing so, have identified multiple genes that simultaneously repress posterior hypothalamic identity while promoting prethalamic identity. This result supports a modified columnar model of organization for the diencephalon, where prethalamus and hypothalamus are situated in adjacent dorsal and ventral domains of the anterior diencephalon. These data serve as a resource for further studies of hypothalamic development, physiology and dysfunction.

Introduction:

The hypothalamus is comprised of a diverse array of neuronal and glial cell types, many of which are organized into spatially discrete clusters or nuclei (*1–3*). Stereotactic lesion and focal stimulation have identified individual nuclei as essential for regulating a broad range of homeostatic physiological processes, ranging from circadian rhythms to hunger; behaviors such as mating, aggression and care of young; and cognitive processes such as motivation, reward, and memory (*4–7*). More recently, opto- and chemogenetic techniques have made progress in identifying the role of individual hypothalamic neuronal subtypes in controlling some of these behaviors (*8–10*).

Progress in this area has been hampered, however, by the fact that hypothalamic cell types thus far have been quite poorly characterized, despite recent efforts aimed at using scRNA-Seq to classify cells in certain hypothalamic regions (*11–14*). Still less is known about how hypothalamic cell types acquire their identities during development. Even the basic spatial organization of the

developing hypothalamus, and its relationship to other forebrain structures such as the prethalamus and telencephalon, are unclear (*15–17*). Previous efforts using microarray analysis coupled with large-scale two-color *in situ* hybridization have identified a set of molecular markers that uniquely define spatial domains of the early embryonic hypothalamus and adjacent diencephalic regions (*2*), while parallel efforts using high-throughput *in situ* hybridization have identified additional region-specific markers (*18*, *19*).

These datasets have been used as the basis for genetic studies that selectively disrupt development of specific hypothalamic regions and/or cell types (*20–24*), leading in some cases to identification of novel functions for previously characterized hypothalamic regions or cell types (*25*, *26*). However, these datasets have important limitations: they do not provide cellular resolution of gene expression data, and they do not efficiently measure coexpression of multiple genes. Despite the availability of many highly specific molecular markers, analysis of mutants that affect hypothalamic development is currently both slow and difficult, owing to the complexity of this structure.

Recent advances in single-cell RNA-Seq technology (scRNA-Seq)(*27*) have made it possible to both analyze development of complex organs at cellular resolution and rapidly and comprehensively characterize the molecular phenotype of developmental mutants (*28*). In this study, we use scRNA-Seq to profile changes in gene expression and cell composition across the full course of mouse hypothalamic development, focusing on early stages during which regionalization and neurogenesis occur, and to rapidly and comprehensively identify changes in hypothalamic patterning in multiple mutant lines. This information reveals common molecular mechanisms controlling patterning of hypothalamus and prethalamus, which is inconsistent with prosomere models of forebrain development (*29*), and is instead consistent with a modified columnar model of diencephalic organization (*15–17*).

Results:

To profile changes in gene expression across the full course of mouse hypothalamic development, we processed 12 timepoints ranging from embryonic day (E)10 to postnatal day (P)45. For E10-E16, both prethalamus and hypothalamus were collected, whereas for E18-P45, only hypothalamus was profiled (Fig. 1a). In total, 129,151 cells were profiled (Fig. S1a,b). Using molecular markers of known hypothalamic regions and cell types (*2*), we were able to annotate all major brain regions and cell types at each individual age (Fig. 1, S1c,d). Aggregation of the entire dataset with UMAP shows a clear developmental progression from neuroepithelial cells at E10 to major cell types in mature hypothalamus (Fig. 1b, Fig. S2a). Reasonably similar detection of expressed genes was observed at each developmental stage (Fig. S2b,c).

Trajectories leading from neuronal progenitors to mature excitatory and inhibitory neurons, glia, ependymal cells and tanycytes were observed, as were separate clusters representing hematopoietic, microglial and endothelial cells (Fig. 1b). Abundant expression of hypothalamic neuropeptides such as *Pomc*,

Agrp, *Ghrh, Sst*, *Gal*, *Hcrt*, and *Pmch* were observed in the neuronal trajectory (Fig. S3).

A separate cluster of oligodendrocytes that does not form a continuous developmental trajectory were observed from P4 onwards, which was marked by high and selective expression of *Glul* (Fig. 1c). Midbrain oligodendrocytes have been recently reported to selectively express *Glul (30)*, potentially indicating an extra-hypothalamic origin of these cells. These data also indicate clear molecular separation between ependymal cells and tanycytes, which are distinguished by molecular markers from E13 onwards, yet imply a common origin for these two ventricular cell populations (Fig. 1d, Fig. S4). Pseudotime analysis with BEAM analysis showed a clear and early bifurcation between ependymal cells and tanycytes, and identified multiple transcription factors that are potential candidates of tanycyte/ependymal differentiation in the hypothalamus (Fig. S4a,b).

We next investigated whether we could use this dataset to resolve spatially distinct domains during early stages of hypothalamic development. To do this, we clustered data from E11, E12, and E13, which correspond to the peak period of hypothalamic neurogenesis (Fig. 2a). Using previously identified region-specific markers as a reference (*2*), we observed a clear segregation of spatially distinct neuronal precursors and progenitors (Fig. 2b, S5). We were able to readily distinguish hypothalamic and non-hypothalamic cell populations, identifying discrete clusters for telencephalic structures such as preoptic area and medial ganglion eminence, thalamic eminence and prethalamus proper, the rim domain and main body of the sensory thalamus, as well as the zona limitans intrathalamica (ZLI)(Fig. 2b, S5).

Each of the previously reported major subdivisions of the developing hypothalamus (*2*) were also identified, including postmitotic neuronal precursor cells of the paraventricular nucleus/supraoptic nucleus (PVN/SON), intrahypothalamic diagonal (ID) and tuberomamillary terminal (TT), ventromedial hypothalamus (VMH), arcuate nucleus (ARC), premamillary hypothalamus (PMN), mamillary nucleus (MMN), and supramamillary nucleus (SMN) (Fig. 2b, S5). In addition, several spatially distinct subtypes of mitotic hypothalamic progenitor cells were also observed, most notably cells that shared markers of both MMN and SMN (Fig. 2b, Table S1).

Multiple known and previously undescribed molecular markers were identified for each of these regions (Fig. 2c, Table S1). While some of these markers are shared among multiple regions in the hypothalamus and other forebrain regions, others are highly specific and non-overlapping. Looking at our dataset in 3D UMAP plot, which robustly maintains global distances during dimensional reduction, we observed a clear separation between multiple forebrain regions, which showed clear dynamic temporal changes in gene expression (Fig. S6). In the hypothalamus, we also identified clear separation between mitotic neural progenitors and postmitotic neural precursors (Fig. 2d), and also observed separate neurogenic trajectories characterized by expression transcription factors that selectively mark posterior (*Pitx2*) and GABAergic (*Arx, Dlx5*) hypothalamic neurons (Fig. 2e, f).

Due to the high complexity of clusters observed in two and three-dimensional tSNE/UMAP plots, it is difficult to comprehensively visualize region-specific differences in gene expression. To improve visualization of these data, we have generated a heatmap for major pattern marker genes that corresponds to the two-dimensional sagittal plane of the developing hypothalamus capturing the main regions (Fig. S7).

This analysis also identified clusters corresponding to three hypothalamic regions that had not been described in previous work (*2*), including two populations of excitatory neurons. The first of these is found in the dorsomedial hypothalamus, and is marked by expression of *Sst*, *Cited1*, *Otp* and *Six6* (Fig. 2c). The second of these is found in the TT/PMN region, and it expresses *Pax7* (Table. S1). The third of these is found in the lateral hypothalamus (LH)consists of a fairly heterogeneous population. This LH cluster consists primarily of glutamatergic neurons, with a small subpopulation of GABAergic neurons (Fig. S8). The glutamatergic population includes a discrete subcluster of *Lhx9*-positive neurons, which marks precursors of hypocretin neurons (*2*, *31*, *32*).

Clustering of previously characterized spatial domains also identified discrete subclusters that express common sets of genes. This is clearly seen in the PVN/SON cluster (Fig. S9). Selective expression of *Onecut2*, *Cartpt* and *Zic1* characterizes a ventrolateral domain that, based on its position, may correspond to the developing SON (Fig. S9). This same approach can be applied to other forebrain regions. We have previously identified molecular markers that both identify discrete spatial domains within the prethalamus and are also shared with hypothalamic neurons (*2*). Subclustering of prethalamic neuronal precursors allows us to detect these and other spatial subdivisions of the prethalamus. We observe partially overlapping expression of the transcription factors *Sp8* and *Sp9* in prethalamus (Fig. S10), which play critical roles in development of interneurons in the telencephalon (*33*). *In situ* hybridization analysis reveals enriched expression of *Sp8* and *Lhx1* in anterior prethalamus and ID, while *Sp9* and *Prox1* are enriched in posterior prethalamus (Fig. S10). *Zic1* and *Ebf*1 also mark discrete spatial domains in developing prethalamus (Fig. S10).

This analysis was able to efficiently identify gene expression patterns restricted to specific spatial domains and subdomains of the developing hypothalamus and prethalamus, confirming and extending our previous findings (*2*). However, this approach does not allow us to identify genes with more complex and/or temporally distinct expression patterns, but which nevertheless play important roles in regulating hypothalamic neurogenesis. To address this, we used scCoGAPS, a non-negative matrix factorization tool that allows unbiased identification of patterns of co-expressed genes (*34*) (Fig. S11). Using this method, we were able to identify patterns that not only matched key spatial subdivisions of the hypothalamus and prethalamus, but also patterns that labeled discrete subsets of hypothalamic progenitors (Table S2). These include NPCs that likely represent radial glia (*Fabp7*, *Slc1a3*), as well as neurogenic progenitors (*Pitx2*, *Nhlh1*, *Nhlh2*). Most strikingly, we observe multiple patterns that selectively label neurogenic progenitors located along the border between

the hypothalamus and prethalamus and both the telencephalon and ZLI (Fig. 2e,f). Genes that drive this pattern include *Neurog2*, *Lhx5*, and *Nhlh2*. Although these expression patterns have been previously reported (*2*), the fate of these border cells is unknown.

These data provide a high-resolution molecular atlas of the developing hypothalamus and prethalamus. We next sought to determine if this approach could be used to rapidly and comprehensively characterize mutants that regulate early stages of hypothalamic development and organization, and which display complex molecular phenotypes. As proof of concept, we performed scRNA-Seq analysis on E12.5 *Foxd1*$^{CreGFP/+}$*;Ctnnb1*$^{ex3/+}$ mice, in which a constitutively active form of beta-catenin is overexpressed in *Foxd1*-positive hypothalamic and prethalamic progenitors, leading to activation of canonical Wnt signaling in these cells and their descendants, along with littermate controls (*20*), along with littermate controls. These mice show broad activation of the canonical Wnt pathway effector *Lef1*, a hyperplastic ventricular zone, and with the exception of a handful of posterior hypothalamic markers, the loss of most regional markers in hypothalamus and prethalamus (*20*).

ScRNA-Seq analysis of E12.5 control and mutant animals reveals changes in gene expression and cell composition that match previously reported findings (Fig. 3a, S12). We observed a substantial increase in undifferentiated NPCs, along with a corresponding reduction in the number of cells expressing markers of hypothalamic and prethalamic neuronal precursors (Fig. 3b, S12c). This was particularly notable for markers shared between hypothalamus and prethalamus, such as *Arx*, *Isl1* and *Gad1/2* (Fig. S13). We identified two cell clusters that are found exclusively in mutant mice, both of which express NPC markers and show high expression of both *Lef1* and negative regulators of canonical Wnt signaling such as *Dkk1*, *Wif1* and *Axin2* (Fig. 3c). One of these two clusters is strongly enriched for G2/M phase markers such *Ube2c*, *Rrm2,* and *Ccnb1* (Fig. 3d), implying that these cells are undergoing mitotic arrest or delay at this stage, as has been previously reported in non-neuronal cells that show high levels of canonical Wnt signaling (*20*, *35*). This finding helps explain the previous observation that, although a massive increase in the number of NPC cells is seen in these mutants, only a modest increase is observed in EdU labeling, which marks NPCs in S-phase (*20*). This demonstrates the power of scRNA-Seq to rapidly and accurately analyze developmental phenotypes in a small fraction of the time and far more comprehensively than is possible using conventional histological techniques.

We next used this same approach to characterize E12.5 *Nkx2.1*$^{CreER/CreER}$ knock-in mice, which are homozygous for a null allele in the homeodomain transcription factor *Nkx2.1* (*36*). *Nkx2.1* is broadly and selectively expressed in ventral hypothalamic progenitors, as well as in progenitors that give rise to telencephalic interneurons (*37*, *38*). Loss of function of *Nkx2.1* leads to a substantial reduction in ventral hypothalamic structures by E18 (*39*), although a detailed characterization of the molecular phenotype of these mutants has not been conducted. Analysis of *Nkx2.1*$^{CreER/CreER}$ and heterozygous littermate controls revealed a broad loss of markers specific to *Nkx2.1*-positive ventral

hypothalamic structures such as ARC, VMH, PMN, and MMN (Fig. 4a,b,c, Fig. S14,15), with both the relative expression levels and number of cells expressing these markers substantially reduced. Ventricular thickness was also reduced in the mutant line (Fig. S16). *Nkx2.1*$^{CreER/CreER}$ also showed an increase in the proportion of prethalamic cells in our scRNA-Seq dataset, in parallel with the loss of ventral hypothalamic structures (Fig. 4b,c, Fig. S17). *In situ* hybridization confirmed that there was an increase in the size of prethalamus and its proportion in the diencephalon (Fig. S17). Ventral hypothalamic regions that are strongly affected by *Nkx2.1* loss of function show changes in cell composition and/or identity (Fig. S18). These structures include ARC, VMH, PMN, and MMN, but not SMN. *Nkx2.1*-negative hypothalamic regions such as the PVN/SON are unaffected. An increase in the number of cells expressing markers of NPCs in the SMN and MMN was also seen (Fig. 4b, S18).

We and others have previously observed extensive similarities between the gene expression patterns and gene regulatory networks controlling development of neuronal progenitors and precursors in the prethalamus and in adjacent hypothalamic regions, such as the ID and TT (*2*, *40*, *41*). This raises the question of whether or not these structures are coordinately patterned. To directly investigate the effects of disrupting genes that might coordinately pattern both prethalamus and hypothalamus, we analyzed E12.5 *Foxd1-Cre;Dlx1/2*$^{lox/lox}$ mice (*42*). *Dlx1* and *Dlx2* are essential for migration and maturation of telencephalic interneuron precursors, and they are also selectively expressed in prethalamus, ID and TT (*2*, *43*). Analysis of *Dlx1/2* mutants and heterozygous littermate controls revealed substantial changes in cell composition and gene expression patterns. We observed a broad reduction in expression levels and the number of cells expressing prethalamic markers such as *Meis2* (Fig. 5a,b,c, Fig. S19, S20). Conversely, a substantial increase in the fraction of cells expressing markers specific to SMN/MMN progenitors such as *Pitx2* and *Nr4a2* was also observed (Fig. 5b,c), implying an expansion of posterior hypothalamic progenitor markers at the expense of prethalamic and ID/TT markers (Fig. S21-S24).

These data strongly suggest that *Dlx1/2* may coordinately regulate patterning and differentiation of prethalamus and hypothalamus, and that both promote specification of prethalamus and ID/TT while repressing posterior hypothalamic identity. To further investigate this possibility that common molecular mechanisms may control patterning of both prethalamus and posterior hypothalamus, we analyzed E12.5 *Nkx2.2*$^{Cre/Cre}$ knock-in mice (*44*). *Nkx2.2* is selectively expressed in NPCs of the posterior prethalamus, as well as in dorsal hypothalamus (*2*, *45*)(Fig. 6). Analysis of *Nkx2.2* mutants and heterozygous littermate controls revealed that the mutants showed a phenotype similar to that of *Dlx1/2* mutants, although more severe (Fig. 6a,b,c; Fig. S25-27). A near-complete loss of prethalamic and ID/TT markers was seen, along with a parallel expansion in the relative fraction of SMN/MMN/prethalamic hybrid progenitors (Fig. 6d). Interestingly, a population of cells is seen in mutant animals that co-express the prethalamic marker *Sp8* along with the SMN/MN progenitor markers *Pitx2* and *Otp* (Fig. S28-30). These cells are absent from controls and highlight

the severe and coordinated disruptions in prethalamic and hypothalamic patterning observed in these mutants (Fig. S28-30). This indicates that *Dlx1/2* and *Nkx2.2* both activate expression of developing prethalamic neurons and repress expression of genes specific to posterior hypothalamic progenitors, highlighting the finding that prethalamus and posterior hypothalamus are coordinately specified and patterned.

Discussion:

In this study, we use scRNA-Seq to develop a molecular atlas of the developing mouse hypothalamus, with a particular focus on stages when hypothalamic patterning and neurogenesis are regulated. This dataset identifies genes that are selectively expressed during the differentiation of major neuronal and non-neuronal hypothalamic cell types, and it accurately delineates spatial subdivisions present in the early stages of hypothalamic development and of the adjacent prethalamic development. This dataset allows rapid and accurate phenotyping of mutants that show broad effects on hypothalamic patterning, neurogenesis and differentiation, in which we were able to validate these findings with traditional histological analysis. Despite the availability of highly specific molecular markers for the major spatial subdivisions of the hypothalamus, the highly complex and temporally dynamic anatomy of this brain region makes analysis of mutant phenotypes slow and complex. Previously, it has taken up to several years of full-time labor to obtain detailed characterization of individual mutant lines. This dataset allows these analyses to be conducted far more rapidly, efficiently, and comprehensively.

Our data suggest that common molecular mechanisms are used in patterning the prethalamus and hypothalamus. Loss of function of both *Dlx1/2* and *Nkx2.2* result in a selective and global loss of expression of molecular markers that are specific to prethalamus and also hypothalamic GABAergic neuronal subpopulations in the ID and TT. We observed a compensatory expansion of posterior hypothalamic progenitor populations, along with the emergence of cells that co-express markers of both prethalamus and posterior hypothalamus. This parallels phenotypes previously observed in anterior hypothalamus, where *Lhx2* loss of function leads to loss of expression of eye field and anterioventral hypothalamic markers at the expense of genes specific to anterodorsal hypothalamic and thalamic eminence (*46*).

These findings cast new light on the topic of the organization of the developing hypothalamus, which has remained highly controversial for decades. The classic columnar model and the more recent prosomere model even differ on the basic position and orientation of the hypothalamus relative to other forebrain structures such as the telencephalon and prethalamus (*15–17*). Our identification of *Dlx1/2* and *Nkx2.2* as simultaneously promoting prethalamic differentiation while also repressing posterior hypothalamic identity supports a modified columnar model of forebrain organization, wherein prethalamus and hypothalamus are situated in adjacent dorsal and ventral domains of the anterior diencephalon (Fig. S31). More generally though, scRNA-Seq analysis of the sort described here makes it possible to definitively determine the organization of

complex structures in the developing brain. All previous models of hypothalamic development were constructed using very sparse datasets - typically single color in situ hybridization of limited number of genes at a small number of timepoints. ScRNA-Seq, which comprehensively catalogs both individual cell types and their gene expression profiles, offers a much richer dataset for constructing such models.

Materials and Methods:

Mice

All experimental animal procedures were approved by the Johns Hopkins University Institutional Animal Care and Use Committee. All mice were housed in a climate-controlled facility (14-hour dark and 10-hour light cycle) with *ad libitum* access to food and water. Time-mated CD1 or C57BL/6J mice were ordered from Charles River Laboratories to collect mice embryos at E10 (CD1), E11 (CD1), E12 (CD1), E13 (CD1), E14 (CD1), E15 (CD1), E16 (1 CD1 and 1 C56BL/6J), and E18 (CD1). Time-mated CD1 mice were ordered to collect pups at P4, P8 and P14. C57BL/6J mice were used for P45 age collection.

Nkx2.1^{CreERT2} knock-in (*36*)(JAX #014552), *Nkx2.2*^{CreEGFP} knock-in (*36*, *44*) (JAX #026880), *Foxd1*^{CreGFP} knock-in (*47*)(JAX #012463), *Ctnnb1*$^{ex3/ex3\ (48)}$, and *Dlx1/2*$^{lox/lox\ (49)}$(JAX #025612) were used for single-cell phenotyping studies. Ctnnb1$^{ex3/ex3}$ mice were crossed with *Foxd1*$^{CreGFP/+}$ to generate *Foxd1*$^{CreGFP/+}$ or *Foxd1*$^{CreGFP/+}$*;Ctnnb1*$^{ex3/+}$. *Dlx1/2*$^{lox/lox}$ mice were crossed with *Foxd1*$^{CreGFP/+}$ to generate *Foxd1*$^{CreGFP/+}$*;Dlx1/2*$^{lox/+}$ or *Foxd1*$^{CreGFP/+}$*;Dlx1/2*$^{lox/lox}$. *Nkx2.1*^{CreERT2}, *Nkx2.2*^{CreEGFP} were C56BL/6J background. *Foxd1*$^{CreGFP/+}$ knock-in, *Ctnnb1*$^{ex3/ex3}$, and *Dlx1/2*$^{lox/lox}$ C57BL/6 and CD1 mixed background. Mice were time-mated during their estrous cycle. E12.5 embryos were collected for generating scRNA-Seq dataset, and E12.5 or E13.5 embryos were collected for histological analysis.

Dissection and cell dissociation

Embryos or postnatal mice were collected and dissociated following previously published protocol (*50*). Embryos were collected using Hibernate-E media (Thermo Fisher Scientific) with 2% B-27 supplement (Thermo Fisher Scientific) and GlutaMAX supplement (0.5 mM final, Thermo Fisher Scientific). A small incision was made dorsal to the lower jaw to expose the ventral portion of the brain. For samples collected between E10 and E16, tissue residing posterior to the medial ganglionic eminence and anterior to the midbrain and sensory thalamus was dissected to collect developing prethalamus and hypothalamus consistently between dissections and when performing phenotyping analysis. Prethalamus was excluded from samples aged E18 and older, with only hypothalamus collected, as previously described (*2*). Collection of medial ganglionic eminence (anterior to the hypothalamus) and posterior structure to the supramamillary nucleus ensured that the equivalent diencephalon area (hypothalamus and prethalamus) was always included. Between 8 and 12 embryos of either sex were collected for each embryonic time point.

Postnatal mice were collected using Hibernate-A media with 2% B-27 and GlutaMAX (0.5 mM final), and the region that is posterior to the optic chiasm and anterior to the posterior hypothalamus were collected. Eight pups (4 male and 4 female) were collected for P4, P8, and P14 dataset, and 3 male mice were pooled for P45 dataset. E10, E12, and E15; E11 and E13; E14, E16, and P45; E18, P4, and P14 were each generated on the same day.

For single-cell phenotyping studies, E12.5 time-mated mice were collected and placed in buffer mentioned above on ice. Tail-tips were collected and rapidly genotyped using GeneAmp Fast PCR mastermix (Thermo Fisher Scientific). Both control and mutant groups were collected on the same day, and E12.5 embryos were pooled from 3 different dams.

Following dissection, tissues were dissociated in papain (Worthington Biochemical) as previously described in calcium-free Hibernate media (*50*). Tissue debris was removed using OptiPrep density gradient media (Sigma-Aldrich) in postnatal mice following cell dissociation. Numbers of viable cells was counted manually via haemocytometer with Trypan Blue staining, and cell concentration was adjusted following manufacturer's protocol of 10x Genomics.

ScRNA-Seq library generation and data processing

Suspended cells were loaded into 10x Genomics Chromium Single Cell System (10X Genomics), and libraries were generated using v1 (1 library) and V2 chemistry with manufacturer's instructions. Libraries were sequenced on Illumina MiSeq (1 library) and NextSeq500. Sequencing data were pre-processed through the Cell Ranger pipeline (10x Genomics) with default parameters, aligned to mm10 genome, and matrix files were used for subsequent bioinformatic analysis.

Data analysis

For analysis of the entire hypothalamic dataset, Seurat V2 (*50*, *51*) and Scanpy (*52*) were used to process matrix files. Datasets were normalized to transcript copies per 10,000, and log-normalized. Top variable genes (4,000 genes) were used for Principal Component Analysis (PCA), and 100 PCA variables were used for tSNE or UMAP dimensional reduction (*53*). Log-normalized values with xyz coordinates from UMAP or xy coordinates from tSNE were extracted and plots were drawn using custom-written script. Individual ages were clustered using Louvain clustering algorithm provided in Scanpy.

For analysis of hypothalamic patterning, the E11, E12, and E13 datasets were extracted from the entire hypothalamic dataset to perform detailed analysis on hypothalamic patterning (*2*) and processed as described above. Initial clustering was using the Louvain clustering algorithm (*54*), and individual clusters were further subdivided to capture all the main regions in the developing diencephalon and nearby structures (*2*). Following sub-divisions, enriched genes in the individual cluster were extracted and cross-referenced to our previous work (*2*), *in situ* hybridization (ISH) validation, Allen Brain Atlas (ABA) (*18*, *19*) and GenePaint to validate our cluster assignments (*18*, *19*).

For mutant phenotyping, individual lines (both control and mutants) were clustered together. Individual regions of the developing diencephalon and nearby brain structures were compared based on the E11-E13 scRNA-Seq dataset, and gene expression focusing on pattern-specific markers between control and mutants were compared. The percentage of regions occupied by either genotype was compared as well. Both changes in pattern-specific markers and percentage of clusters occupied by each genotype reflect biological changes, but not variance that occurred during dissection and library preparation.

scCoGAPS analysis

scCoGAPS, a Bayesian non-negative matrix factorization algorithm, was used to identify 'patterns' in E11-E13 dataset as described previously (*55*), except using all genes to identify patterns. Up to 60 patterns were identified and named by projecting scCoGAPS patterns into our 3D UMAP plot and using our prior knowledge of gene expression patterns in the developing hypothalamus. Patterns obtained from scCoGAPS were used for pseudotime analysis by subsetting cells that belong to the particular pattern.

E12 spatial mapping

XY coordinates of E12 were drawn based on our previous work to capture all post-mitotic regions of the developing diencephalon, average values of individual clusters were assigned to each XY coordinates, and 2D spatial distribution was drawn using SpatialDE (*34*, *56*).

Cell cycle analysis

Cell cycling from *Ctnnb1*-overexpressing cluster of mutant line *Foxd1*$^{Cre/+}$;*Ctnnb1*$^{Ex3/+}$ were analyzed using scran (*57*).

Pseudotime analysis

Monocle2 (*58*) was used to perform pseudotime analysis to understand differences in developing of tanycytes and ependymal cells; patterns that we have obtained from scCoGAPS.

***In situ* hybridization (ISH)**

Chromogenic *in situ* hybridization was performed as previously described (*2*), except E12.5 or E13.5 embryos were fixed with 4% paraformaldehyde, sectioned at 25 um with either coronal or sagittal plane and treated with proteinase K for 5 min at room temperature. Single-molecule fluorescent *in situ* hybridization was done using RNAScope with probes targeting *Meis2*, *Sp8*, *Sp9*, *Irx5*, *Pitx2*, and *Lhx6* on E12.5 or E13.5 embryos following the manufacturer's protocol.

EdU

EdU (20 mg/ml) was injected 3 hours prior to collection into time-mated dams and embryos were collected and processed, as previously described (*2*, *20*).

Immunostaining

Fixed embryos were processed for immunostaining with Pax6-antibody (1:200, AB2237, EMN Millipore). Sections were mounted with Vectamount (Vectorlabs).

Contribution: DWK, SB designed experiments. DWK, PWW, ZQW, SL, LJ performed experiments. DWK, PWW, ZQW, SL, CS analyzed data,

Acknowledgements:
This work was supported by a grant from the NIH (DK108230) to S.B. Transcriptomics and Deep Sequencing Core at Johns Hopkins for sequencing all scRNA-Seq libraries. *Dlx1/2*$^{lox/lox}$ mice were provided by Jae Lee (OSHU). We thank M. Placzek, E. Newman, J. Nathans, A. Kolodkin, W. Yap, and members of the Blackshaw lab for comments on the manuscript.

Data availability:
All scRNA-Seq data are available on GEO. Data can be viewed at https://proteinpaint.stjude.org/F/mm10/example.scrna.html.

Figure legends:
Figure 1. Overview of generation of the hypothalamus scRNA-Seq dataset. (a) Schematic diagram showing overall experimental strategy. 12 timepoints of developing diencephalon including prethalamus and hypothalamus (between E10 and E16), and hypothalamus (between E18 and P45) were profiled using 10x Genomics Chromium system. Distribution of individual ages is shown in 3D UMAP plot. (b) 3D UMAP plot showing distribution of major cell types (red) of the hypothalamus at the terminal neuronal branch. (c) 3D UMAP plot (higher power view) showing two separate populations of oligodendrocytes, *Glul*-positive and *Glul*-negative oligodendrocytes in the hypothalamus. (d) 3D UMAP plot (higher power view) showing developmental bifurcation leading to development of ependymal cells (green, *Hdc*) and specialized-ependymal cells known as tanycytes (brown, *Rax*, *Cd44*).

Figure 2. Specification of hypothalamic patterning during embryonic development. (a) Schematic diagram showing extraction of E11, E12 and E13 data to perform detailed analysis on hypothalamic patterning during development. (b) 2D tSNE plot showing divisions of developing diencephalon (prethalamus and hypothalamus) and nearby regions to the developing diencephalon. (c) Heatmap showing pattern-specific genes in individual main hypothalamic regions. (d) scCoGAPS analysis capturing hypothalamic NPC (blue) and neurons (red)(top) and NPC to neurogenic differentiation (bottom). (e) Pseudotime analysis on scCoGAPS capturing hypothalamic NPC to neurogenic differentiation and (f) Pseudotime-heatmap from Monocle2 BEAM analysis.

Figure 3. scRNA-Seq can be used to screen mutant phenotype in the hypothalamus. (a) 2D tSNE showing distribution of diencephalic and nearby brain structures. (b) 2D tSNE heatmap showing distribution of individual cluster between control (*Foxd1*$^{Cre/+}$) and hypothalamic *Ctnnb1* over-expressing mutant line (*Foxd1*$^{Cre/+}$*;Ctnnb1*$^{Ex3/+}$). Note distinct clusters that only exist in the mutant group (mutant-specific clusters). (c) 2D tSNE plot showing high level of *Lef1*, *Rrm2*, *Wif1* and *Axin2* in the mutant line. (d) Heatmap showing that mutant-specific clusters display all pattern-specific markers compare to the entire control hypothalamic population, with a high level of G2/M phase markers.

Figure 4. Loss of ventral hypothalamic structures in *Nkx2.1* mutant line. (a) 2D tSNE showing distribution of diencephalic and nearby brain structures. (b) 2D tSNE heatmap showing distribution of individual cluster between control (*Nkx2.1*$^{CreER/+}$) and *Nkx2.1* mutant line (*Nkx2.1*$^{CreER/ CreER}$). Note absence of ventral diencephalic structures (except the supramamillary nucleus), and expansion and higher proportion of prethalamus in the mutant line. (c) 2D tSNE plot showing low level of *Chchd10*, *Hmx2* and *Foxb1* in the ventral hypothalamic structures, and high level of *Sp8* (prethalamus) in the mutant line. (d) Schematic diagram and in situ hybridization of *Nkx2.1* distribution (orange = ventral diencephalon) and schematic summary of loss of Nkx2.1 phenotype in the diencephalon.

Figure 5. Reduction of prethalamic structures and expansion of posterior-ventral hypothalamus is seen in *Dlx1/2* mutants. (a) 2D tSNE showing distribution of diencephalic and nearby brain structures. (b) 2D tSNE heatmap showing distribution of individual cluster between control (*Foxd1-Cre;Dlx1/2*$^{lox/+}$) and *Dlx1/2* conditional knockout line (*Foxd1-Cre;Dlx1/2*$^{lox/lox}$). Note absence of dorsal diencephalic structures (prethalamus) and expansion of the mammillary nuclei (medial mamillary and supramamillary) in the mutant line. (c) 2D tSNE plot showing low level of *Gad1* and *Meis2* (prethalamus), and high level of *Pitx2* and *Nr4a2* (mamillary nuclei) in the mutant line. (d) Schematic diagram and *in situ* hybridization of *Foxd1* and *Dlx1/2* distribution and schematic summary of loss of *Dlx1/2* phenotype in the diencephalon. Note a reduction in the prethalamus and expansion of medial- and supramamillary nucleus.

Figure 6. Reduction of prethalamic structures and rise of hybrid population in *Nkx2.2* mutant line. (a) 2D tSNE showing distribution of diencephalic and nearby brain structures. (b) 2D tSNE heatmap showing distribution of individual cluster between control (*Nkx2.2*$^{Cre/+}$) and *Nkx2.2* mutant line (*Nkx2.2*$^{Cre/Cre}$). Note absence of dorsal diencephalic structures (prethalamus) and rise of hybrid population that expresses both prethalamic and mammillary nuclei markers in the mutant line. (c) 2D tSNE plot showing *Otp* and *Pitx2* (mammillary nuclei) in hybrid population and low level of *Sp8* (prethalamus) and slight increase in *Hes1* in the mutant line. (d) Schematic diagram and *in situ* hybridization of *Nkx2.2* distribution and schematic summary of loss of *Nkx2.2* phenotype in the diencephalon.

Extended Figure 1. Overview of individual time points collected for scRNA-Seq analysis. (a) Table showing number of cells collected for scRNA-Seq dataset. (b) 2D tSNE showing distribution of individual ages. (c) 2D tSNE showing distribution of individual prethalamic and hypothalamic regions (nuclei) between E10 and E15. (d) 2D tSNE showing distribution of individual major cell types of the hypothalamus between E16 and P45.

Extended Figure 2. Overview of individual ages in 3D UMAP plot. (a) 3D UMAP plot showing distribution of individual ages collected for scRNA-Seq. (b) Violin graph showing distribution of number and mean (red dot) of genes in individual scRNA-Seq libraries. (c) Violin graph showing distribution of number and mean (red dot) of total mRNAs (UMI) in individual scRNA-Seq libraries.

Extended Figure 3. Distribution of hypothalamic neuronal cell types. (a) 3D UMAP plot highlighting the terminal trajectory of hypothalamic neurons. (b-e) 3D UMAP plot highlighting expression of *Pomc* (b), *Kiss1* (c), *Agrp* (d), *Npy* (e). (f) 3D UMAP plot highlighting expression of *Slc17a6*. (g-j) 3D UMAP plot highlighting expression of *Gnrh* (g), *Ghrh* (h), *Npvf* (i), *Vip* (j). (k) 3D UMAP plot highlighting expression of *Slc32a1*. (l-o) 3D UMAP plot highlighting expression of *Sst* (l), *Gal* (m), *Hcrt* (n), *Pmch* (o).

Extended Figure 4. Pseudotime analysis on differentiation of common precursors into ependymal cells or tanycytes. (a) Pseudotime-trajectory from pre-branch (blue) to differentiation into ependymal-lineage (green) or tanycyte-lineage (red). (b) Pseudotime heatmap from Monocle2 BEAM analysis focusing on transcription factors that are involved with differentiation of a common progenitor into ependymal cells or tanycytes.

Extended Figure 5. 2D tSNE plot showing divisions of developing diencephalon (prethalamus and hypothalamus) and nearby regions to the developing diencephalon.

Extended Figure 6. 3D UMAP plot of developing hypothalamus. (a) 3D UMAP plot showing divisions of developing diencephalon (prethalamus and hypothalamus) and nearby regions to the developing diencephalon. (b-g) 3D UMAP plot showing gene expressions – *Ccnd1* (blue) and *Ccnb1* (red)(b), *Ccnd1* (blue) and *Tubb3* (red)(c); *Chchd10* (blue, arcuate nucleus) and *Otp* (red, PVN/SON)(d), *Pitx2* (blue, SMN) and *Hmx2* (red, PMN)(e); *Pkib* (blue, anterior neural progenitor) and *Gata3* (red, rim domain)(f), *Meis2* (blue, prethalamus) and *Tbr1* (red, thalamic eminence)(g).

Extended Figure 7. 2D sagittal atlas of developing diencephalon (hypothalamus and prethalamus) and projection of pattern-specific genes demarcating individual diencephalic regions.

Extended Figure 8. Detailed analysis of LH shows GABAergic and non-GABAergic clusters. (a) Schematic highlighting LH region. (b-d)(g-j) 2D tSNE plot showing clusters (b), *Sp9* (c), *Dlx2* (d), *Gad2* (g), *Arx* (h), *Nhlh2* (i), *Lhx9* (j). (e-f)(k-l) *In situ* hybridization showing *Dlx1* (e), *Sp9* (f), *Nhlh2* (k), *Gad2* (l).

Extended Figure 9. Detailed analysis of PVN and SON during embryonic development. (a) Schematic highlighting PVN/SON region. (b-d)(h-k)(o-r) 2D tSNE plot showing clusters (b), *Onecut2* (c), *D930028M14Rik* (d), *Cartpt* (h), *Cbln1* (i), *Zic or Zic3* (j), *Fezf1* (k), *Calcr* (o), *Neurod6* (p), *Nkx2-1* (q), *Pou3f2* (r). (e-g)(l-n)(s-u) *In situ* hybridization showing *D930028M14Rik* (e), *Fezf1* (f), *Onecut2* (g), *Cartpt* (l), *Zic1* (m), *Zic3* (n), *Cartpt* – coronal plane (s), *Zic1* – coronal plane (t), *Zic3* – coronal plane (u).

Extended Figure 10. Detailed analysis of prethalamus. (a) Schematic highlighting prethalamic region. (b-d)(g-j) 2D tSNE plot showing clusters (b), *Sp8* (c), *Sp9* (d), *Prox1* (g), *Lhx1* (h), *Zic1* (i), *Ebf1* (j). (e-f)(k-l) *In situ* hybridization showing *Prox1* (e), *Lhx1* (f), *Zic1* (k), *Ebf1* (l). (m-q) RNAScope shows *Sp9*-positive (n) or *Sp8*-positive regions (o) in the prethalamus, and regions that express both *Sp8* and *Sp9* (p, q).

Extended Figure 11. scCoGAPS analysis on developing hypothalamus reveals patterns capturing development or specific subregions of the hypothalamus. scCoGAPS patterns projected into 3D UMAP plot (shown in different angles) of the developing hypothalamus capturing – hypothalamic NPC (a), Hypothalamic NPC G/M phase (b); Hypothalamic NPC subset (c), early neurogenic prethalamic and MGE lineage (d); hypothalamic neurogenic boarder zone (e), PVN & SMN & MMN (f); PVH & VMH & SMN (g), GABAergic populations of MGE & LH (h); GABAergic populations of MGE & vAH & LH (i), Hypothalamic NPC & neurogenic populations (j); hypothalamic neurons (k), PVN & VMH (l); LH & VMH (m), AH NPC & PVH (n); thalamic progenitor (o), thalamic eminence (p); PVN subset (q), PVN subset (r); prethalamus (s), VMH subset (t); VMH subset (u).

Extended Figure 12. Overview of individual groups in *Ctnnb1* line. (a) 2D tSNE showing distribution of cells in individual scRNA-Seq libraries. (b) Violin graph showing distribution of number and mean (black dot) of genes in individual scRNA-Seq libraries. (c) Violin graph showing distribution of number and mean (black dot) of total mRNAs (UMI) in individual scRNA-Seq libraries. (d) Bar graph showing distribution of individual scRNA-Seq libraries.

Extended Figure 13. Heatmap showing differential gene expression between control and *Ctnnb1* mutant line. Heatmap showing top different genes between control and *Ctnnb1*-mutant line in different hypothalamic regions.

Extended Figure 14. Overview of individual groups in *Nkx2.1* line. (a) 2D tSNE showing distribution of cells in individual scRNA-Seq libraries. (b) Violin

graph showing distribution of number and mean (black dot) of genes in individual scRNA-Seq libraries. (c) Violin graph showing distribution of number and mean (black dot) of total mRNAs (UMI) in individual scRNA-Seq libraries. (d) Bar graph showing distribution of individual scRNA-Seq libraries.

Extended Figure 15. Heatmap showing differential gene expression between controls and *Nkx2.1* mutants. Heatmap showing top different genes between control and *Nkx2.1*-mutant line in different hypothalamic regions.

Extended Figure 16. Characterization of *Nkx2.1* mutants. (a) 2D tSNE heatmap showing distribution of individual cluster between control (*Nkx2.1*$^{CreER/+}$) and *Nkx2.1* mutant line (*Nkx2.1*$^{CreER/ CreER}$). Note absence of ventral diencephalic structures (except supramamillary nucleus) and expansion of prethalamus in the mutant line. (b) 2D tSNE plot showing absence of *Nkx2.1* expression in the mutant line. (c–d) Images showing EdU-positive cells at E13.5 in the control (c) and mutant (d). Magnified images are shown in (c') and (d'). Note reduction of thickness in the ventral portion of the hypothalamus and an increase in EdU staining in the dorsal portion of the hypothalamus.

Extended Figure 17. *Nkx2.1* mutants show an increase in the relative proportion of prethalamic cells. (a–d, g-j) Chromogenic *in situ* hybridization showing *Arx* (a, g), *Dlx5* (b, h), *Gad2* (c, i) and *Isl1* (d, j) in control (a-d) and mutant (g-j). (e-f, k-l) RNAScope showing *Meis2* (e-f, k-l) in control (e-f) and mutant (k-l). (m-p, v-y) 2D tSNE showing *Arx* (m, v), *Dlx5* (n, w), *Gad2* (o, x) and *Isl1* (p, y) in control (m-p) and mutant (v-y). Note an increase in the proportion of prethalamus and ventralization of the dorsal portion of the diencephalon. (q-u, z-c') RNAScope showing *Sp9* (r, a'), *Pitx2* (s, b') and *Sp8* (t, c') in control (q-u) and mutant (z-c').

Extended Figure 18. *Nkx2.1* mutant line shows a decrease in ventral hypothalamic cells, with the exception of the supramamillary nucleus. (a–j) Chromogenic *in situ* hybridization showing *Rax* (a, f), *Sox14* (b, g), *Hmx2* (c, h), *Foxb1* (d, i) and *Foxa1* (e, j) in control (a-e) and mutant (f-j). (k–t) 2D tSNE showing *Rax* (k, p), *Sox14* (l, q), *Hmx2* (m, r), *Foxb1* (n, s) and *Foxa1* (o, t) in control (k-o) and mutant (p-t). Note a decrease in ventral hypothalamic markers including arcuate nucleus, ventromedial hypothalamus, premamillary nucleus and medial mammillary nucleus. Supramamillary nucleus shows no change in overall size, but some ectopic expression of supramamillary markers are observed.

Extended Figure 19. Overview of individual groups in *Dlx1/2* line. (a) 2D tSNE showing distribution of cells in individual scRNA-Seq libraries. (b) Violin graph showing distribution of number and mean (black dot) of genes in individual scRNA-Seq libraries. (c) Violin graph showing distribution of number and mean (black dot) of total mRNAs (UMI) in individual scRNA-Seq libraries. (d) Bar graph showing distribution of individual scRNA-Seq libraries.

Extended Figure 20. Heatmap showing differential gene expression between controls and *Dlx1/2* mutants. Heatmap showing top different genes between controls and *Dlx1/2*-mutants in different hypothalamic regions.

Extended Figure 21. Characterization of *Dlx1/2* mutants. (a) 2D tSNE heatmap showing distribution of individual cluster between control (*Foxd1-Cre;Dlx1/2*$^{lox/+}$) and *Dlx1/2* conditional knockout line (*Foxd1-Cre;Dlx1/2*$^{lox/lox}$). Note absence of dorsal diencephalic structures (prethalamus) and expansion of mammillary nucleus (medial mamillary and supramamillary) in the mutant line. (b) Violin plot showing reduction of *Dlx1* and *Dlx2* expression in the mutant line. (c) *In situ* hybridization showing absence of *Dlx2* expression in the developing diencephalon of the mutant line. Note that *Dlx2* expression remains in the medial ganglionic eminence. (d–e) Images showing EdU-positive cells at E13.5 in the control (d) and mutant (e). Magnified images are shown in (d') and (e'). Note an increase in EdU staining in the ventral-posterior portion of the hypothalamus.

Extended Figure 22. *Dlx1/2* mutants show a reduction in the size of prethalamus and a loss of GABAergic markers in both prethalamus and hypothalamus. (a–h) Chromogenic *in situ* hybridization showing *Arx* (a, e), *Isl1* (b, f), *Gad2* (c, g) and *Dlx5* (d, h) in control (a-d) and mutant (e-h). (i –p) 2D tSNE showing *Arx* (i, m), *Isl1* (j, n), *Gad2* (k, o) and *Dlx5* (l, p) in control (i-l) and mutant (m-p). Note a reduction in the proportion of prethalamus and GABAergic markers.

Extended Figure 23. *Dlx1/2* mutants show a selective increase in the size of medial- and supramamillary nucleus. (a–j) Chromogenic *in situ* hybridization showing *Foxb1* (a, f), *Hmx2* (b, g), *Sox14* (c, h), *Foxa1* (d, i) and *Irx5* (e, j) in control (a-e) and mutant (f - j). (k–t) 2D tSNE showing *Foxb1* (k, p), *Hmx2* (l, q), *Sox14* (m, r), *Foxa1* (n, s) and *Irx5* (o, t) in control (k-o) and mutant (p-t). Note a decrease in *Foxb1*$^{+}$ medial mammillary nucleus. Decrease in premamillary nucleus (*Hmx2*), with an expansion in supramamillary nucleus (*Foxa1*, *Irx5*). *Sox14,* expressed in rim domain (*59*), showed an ectopic expression crossing ZLI with an absence of *Sox14* in ZLI itself.

Extended Figure 24. Selective loss of ID and prethalamus with an increase in the supramamillary nucleus in *Dlx1/2* mutant line. (a-j) RNAScope showing *Irx5* (b, g), *Meis2* (c, h) and *Lhx6* (d, i) in control (a-e) and mutant (f-j). Note an expansion of *Irx5* (SMN), and loss of *Meis2* (ID/prethalamus) and *Lhx6* (ID). Note an absence of Lhx6 expression in mutant. (k-t) RNAScope showing *Sp9* (l, q), *Meis2* (m, r) and *Sp8* (n, s) in control (k-o) and mutant (p-t). Note an expansion of prethalamic markers *Sp9*, *Meis2* and *Sp8* in mutant. (u-x) RNAScope showing *Pitx2* (SMN) in control (u-v) and mutant (w-x). Note an increase in *Pitx2* expression in mutant. (y-b') Immunostaining showing Pax6 (prethalamus) in control (y-z) and mutant (a'-b'). Note a decrease and disruption in Pax6 expression in the prethalamus.

Extended Figure 25. Overview of individual groups in *Nkx2.2* line. (a) 2D tSNE showing distribution of cells in individual scRNA-Seq libraries. (b) Violin graph showing distribution of number and mean (red dot) of genes in individual scRNA-Seq libraries. (c) Violin graph showing distribution of number and mean (red dot) of total mRNAs (UMI) in individual scRNA-Seq libraries. (d) Bar graph showing distribution of individual scRNA-Seq libraries.

Extended Figure 26. Heatmap showing differential gene expression between control and *Nkx2.2* mutant line. Heatmap showing top different genes between control and *Nkx2.2*-mutant line in different hypothalamic regions.

Extended Figure 27. Characterization of *Nkx2.2* mutant line. (a) 2D tSNE heatmap showing distribution of individual cluster between control (*Nkx2.2*$^{Cre/+}$) and *Nkx2.2* mutant line (*Nkx2.2*$^{Cre/Cre}$). Note absence of dorsal diencephalic structures (prethalamus) and rise of hybrid population in the mutant line. (b) 2D tSNE plot showing reduction in *Nkx2.2* expression in the mutant line. (c) Violin plot showing reduction of *Nkx2.2* expression in the mutant line. (d–e) Images showing EdU-positive cells at E13.5 in the control (d) and mutant (e). Magnified images are shown in (d') and (e'). Note an increase in EdU staining in the dorsal portion of the ventral hypothalamus.

Extended Figure 28. *Nkx2.2* mutant line shows a reduction in the size of prethalamus and supramamillary nucleus. (a–d)(g-j) Chromogenic *in situ* hybridization showing *Arx* (a, g), *Dlx5* (b, h), *Gad2* (c, i) and *Foxa1* (d, j) in control (a-d) and mutant (g-j). (e-f)(k-l) Pax6 immunostaining in control (e, f) and mutant (k, l). (m–v) 2D tSNE showing *Arx* (m, r), *Dlx5* (n, s), *Gad2* (o, t), *Foxa1* (p, u), and *Pax6* (q, v) in control (m-q) and mutant (r-v). Note a reduction In the proportion of prethalamus (*Arx, Dlx5, Gad2, Pax6*) and an increase in the proportion of supramamillary cells (*Foax1*).

Extended Figure 29. *Nkx2.2* mutant line shows a hybrid population that express both prethalamic and mamillary markers. (a-e, g-k) Chromogenic *in situ* hybridization showing *Foxb1* (a, g), *Hmx2* (b, h), *Otp* (c, i), *Zic1* (d, j) and *Irx5* (e, k) in control (a-e) and mutant (g-k). (f, l, r, x) RNAScope showing *Lhx6* (ID and TT) in control (f, r) and mutant (l, x). Note a selective loss of *Lhx6* expression in the posterior ID of the mutant. (m-q, s-w) 2D tSNE showing *Foxb1* (m, s), *Hmx2* (n, t), *Otp* (o, u), *Zic1* (p, v) and *Irx5* (q, w) in control (m-q) and mutant (s-w). Note a cluster of cells that express pre-, medial- and supramamillary nucleus markers that expand into the prethalamus. scRNA-Seq shows that these populations also express prethalamic markers (Fig 6, Fig S28).

Extended Figure 30. Hybrid cell populations expressing both supramamillary and prethalamic markers are detected *Nkx2-2* mutant line. (a-e) RNAScope showing *Sp9* (c, prethalamus), *Pitx2* (d, SMN) positive hypothalamic region in mutant. White circles in (e) indicates *Sp9* and *Pitx2*

double-positive cells. (f-j) RNAScope showing *Irx5* (c, SMN), *Meis2* (d, prethalamus) positive hypothalamic region in mutant. White circle in (j) indicates a region that lies between supramamillary nucleus and prethalamus expression both *Irx5* and *Meis2*.

Extended Figure 31. Schematic summarizing updated columnar model for organization of the prethalamus and hypothalamus. (a) An updated columnar model suggesting three different major compartments of the embryonic (E11) developing forebrain – 1. telencephalic region, 2. anterior diencephalic region (hypothalamus, prethalamus and thalamic eminence), 3. posterior diencephalic region, comprising sensory thalamus and protector. Note multiple contiguous domains of transcription factors in both prethalamus and hypothalamus. (b-g) Embryonic (E11) developing forebrain showing expression pattern of *Lhx5* (b), *Isl1* (c), *Dlx1/2* (d), *Lhx2* (e), *Prox1* (f), *Nkx2-2* (g) in the diencephalon spanning prethalamus, thalamic eminence, mamillary bodies and tuberal hypothalamus, adapted from (*2*, *60*). (h) Traditional columnar model and prosomeric model showing anterior-posterior axis of the forebrain, and subdivisions of telencephalon (purple) and diencephalon (yellow). Mutant studies showing – 1. conversion of anterioventral hypothalamus into thalamic eminence and paraventricular nucleus in *Lhx2* mutants, adapted from (*46*) (i), 2. hybrid populations that express both prethalamic, medial- and supramamillary nucleus in *Nkx2.2* mutants in this study (j), 3. Expansion of medial- and supramamillary nucleus into the prethalamic region in *Dlx1/2* mutants in this study.

References:

1. J. L. Bedont, E. A. Newman, S. Blackshaw, Patterning, specification, and differentiation in the developing hypothalamus. *Wiley Interdiscip. Rev. Dev. Biol.* **4**, 445–468 (2015).

2. T. Shimogori *et al.*, A genomic atlas of mouse hypothalamic development. *Nat. Neurosci.* **13**, 767–775 (2010).

3. Y. Xie, R. I. Dorsky, Development of the hypothalamus: conservation, modification and innovation. *Development*. **144**, 1588–1599 (2017).

4. M. A. Kent, R. H. Peters, Effects of ventromedial hypothalamic lesions on hunger-motivated behavior in rats. *J. Comp. Physiol. Psychol.* **83**, 92–97 (1973).

5. M. R. Kruk *et al.*, Discriminant analysis of the localization of aggression-inducing electrode placements in the hypothalamus of male rats. *Brain Research*. **260** (1983), pp. 61–79.

6. J. H. Lammers, M. R. Kruk, W. Meelis, A. M. van der Poel, Hypothalamic substrates for brain stimulation-induced attack, teeth-chattering and social grooming in the rat. *Brain Res.* **449**, 311–327 (1988).

7. G. R. Hervey, The effects of lesions in the hypothalamus in parabiotic rats. *J. Physiol.* **145**, 336–352 (1959).

8. J. L. Bedont *et al.*, An LHX1-Regulated Transcriptional Network Controls Sleep/Wake Coupling and Thermal Resistance of the Central Circadian Clockworks. *Curr. Biol.* **27**, 128–136 (2017).

9. J. Kohl *et al.*, Functional circuit architecture underlying parental behaviour. *Nature*. **556**, 326–331 (2018).

10. T. Yang *et al.*, Social Control of Hypothalamus-Mediated Male Aggression. *Neuron*. **95**, 955–970.e4 (2017).

11. R. A. Romanov *et al.*, Molecular interrogation of hypothalamic organization reveals distinct dopamine neuronal subtypes. *Nat. Neurosci.* **20**, 176–188 (2017).

12. L. E. Mickelsen *et al.*, Single-cell transcriptomic analysis of the lateral hypothalamic area reveals molecularly distinct populations of inhibitory and excitatory neurons. *Nat. Neurosci.* **22**, 642–656 (2019).

13. R. Chen, X. Wu, L. Jiang, Y. Zhang, Single-Cell RNA-Seq Reveals Hypothalamic Cell Diversity. *Cell Reports*. **18** (2017), pp. 3227–3241.

14. J. N. Campbell *et al.*, A molecular census of arcuate hypothalamus and median eminence cell types. *Nat. Neurosci.* **20**, 484–496 (2017).

15. L. W. Swanson, *Brain Maps: Structure of the Rat Brain* (Elsevier Publishing Company, 1992).

16. H. Kuhlenbeck, *The Central Nervous System of Vertebrates: pt. 1. Structural elements: biology of nervous tissue. pt. 2. Overall morphologic pattern* (1967).

17. J. L. Rubenstein, S. Martinez, K. Shimamura, L. Puelles, The embryonic vertebrate forebrain: the prosomeric model. *Science*. **266**, 578–580 (1994).

18. A. Visel, C. Thaller, G. Eichele, GenePaint.org: an atlas of gene expression patterns in the mouse embryo. *Nucleic Acids Res.* **32**, D552 (2004).

19. E. S. Lein *et al.*, Genome-wide atlas of gene expression in the adult mouse brain. *Nature*. **445**, 168–176 (2007).

20. E. A. Newman, D. Wu, M. M. Taketo, J. Zhang, S. Blackshaw, Canonical Wnt signaling regulates patterning, differentiation and nucleogenesis in mouse hypothalamus and prethalamus. *Dev. Biol.* **442**, 236–248 (2018).

21. E. A. Newman *et al.*, Foxd1 is required for terminal differentiation of anterior hypothalamic neuronal subtypes. *Dev. Biol.* **439**, 102–111 (2018).

22. D. M. Kurrasch *et al.*, The neonatal ventromedial hypothalamus transcriptome reveals novel markers with spatially distinct patterning. *J. Neurosci.* **27**, 13624–13634 (2007).

23. B. Lee *et al.*, Dlx1/2 and Otp coordinate the production of hypothalamic GHRH- and AgRP-neurons. *Nat. Commun.* **9**, 2026 (2018).

24. K. Sokolowski *et al.*, Specification of select hypothalamic circuits and innate behaviors by the embryonic patterning gene dbx1. *Neuron*. **86**, 403–416 (2015).

25. K. Liu *et al.*, Lhx6-positive GABA-releasing neurons of the zona incerta promote sleep. *Nature*. **548**, 582–587 (2017).

26. J. L. Bedont *et al.*, Lhx1 controls terminal differentiation and circadian function of the suprachiasmatic nucleus. *Cell Rep.* **7**, 609–622 (2014).

27. T. Stuart, R. Satija, Integrative single-cell analysis. *Nature Reviews Genetics*. **20** (2019), pp. 257–272.

28. B. S. Clark *et al.*, Single-Cell RNA-Seq Analysis of Retinal Development Identifies NFI Factors as Regulating Mitotic Exit and Late-Born Cell Specification. *Neuron* (2019), , doi:10.1016/j.neuron.2019.04.010.

29. L. Puelles, J. L. R. Rubenstein, A new scenario of hypothalamic organization: rationale of new hypotheses introduced in the updated prosomeric model. *Front. Neuroanat.* **9**, 27 (2015).

30. W. Xin *et al.*, Oligodendrocytes Support Neuronal Glutamatergic Transmission via Expression of Glutamine Synthetase. *Cell Rep.* **27**, 2262–2271.e5 (2019).

31. J. Liu *et al.*, Evolutionarily conserved regulation of hypocretin neuron specification by Lhx9. *Development*. **142**, 1113–1124 (2015).

32. J. Dalal *et al.*, Translational profiling of hypocretin neurons identifies candidate

molecules for sleep regulation. *Genes Dev.* **27**, 565–578 (2013).

33. J. Li *et al.*, Transcription Factors Sp8 and Sp9 Coordinately Regulate Olfactory Bulb Interneuron Development. *Cereb. Cortex*. **28**, 3278–3294 (2018).

34. G. L. Stein-O'Brien *et al.*, Decomposing cell identity for transfer learning across cellular measurements, platforms, tissues, and species. *bioRxiv* (2018), p. 395004.

35. D. Olmeda, S. Castel, S. Vilaró, A. Cano, β-Catenin Regulation during the Cell Cycle: Implications in G2/M and Apoptosis. *Molecular Biology of the Cell*. **14** (2003), pp. 2844–2860.

36. H. Taniguchi *et al.*, A resource of Cre driver lines for genetic targeting of GABAergic neurons in cerebral cortex. *Neuron*. **71**, 995–1013 (2011).

37. T. Du, Q. Xu, P. J. Ocbina, S. A. Anderson, NKX2.1 specifies cortical interneuron fate by activating Lhx6. *Development*. **135**, 1559–1567 (2008).

38. L. A. B. Elias, G. B. Potter, A. R. Kriegstein, A time and a place for nkx2-1 in interneuron specification and migration. *Neuron*. **59**, 679–682 (2008).

39. S. Kimura *et al.*, The T/ebp null mouse: thyroid-specific enhancer-binding protein is essential for the organogenesis of the thyroid, lung, ventral forebrain, and pituitary. *Genes Dev.* **10**, 60–69 (1996).

40. C. N. Sunnen, J. C. Simonet, E. D. Marsh, J. A. Golden, Arx is required for specification of the zona incerta and reticular nucleus of the thalamus. *J. Neuropathol. Exp. Neurol.* **73**, 253–261 (2014).

41. N. Moreno, L. Domínguez, S. Rétaux, A. González, Islet1 as a marker of subdivisions and cell types in the developing forebrain of Xenopus. *Neuroscience*. **154**, 1423–1439 (2008).

42. B. A. Seybold *et al.*, Chronic reduction in inhibition reduces receptive field size in mouse auditory cortex. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 13829–13834 (2012).

43. A. Simeone *et al.*, Cloning and characterization of two members of the vertebrate Dlx gene family. *Proc. Natl. Acad. Sci. U. S. A.* **91**, 2250–2254 (1994).

44. D. A. Balderes, M. A. Magnuson, L. Sussel, Nkx2.2:Cre knock-in mouse line: A novel tool for pancreas- and CNS-specific gene deletion. *genesis*. **51** (2013), pp. 844–851.

45. A. Holz *et al.*, The transcription factors Nkx2.2 and Nkx2.9 play a novel role in floor plate development and commissural axon guidance. *Development*. **137**, 4249–4260 (2010).

46. A. Roy *et al.*, LHX2 is necessary for the maintenance of optic identity and for the progression of optic morphogenesis. *J. Neurosci.* **33**, 6877–6884 (2013).

47. B. D. Humphreys *et al.*, Fate tracing reveals the pericyte and not epithelial origin of myofibroblasts in kidney fibrosis. *Am. J. Pathol.* **176**, 85–97 (2010).

48. N. Harada *et al.*, Intestinal polyposis in mice with a dominant stable mutation of the beta-catenin gene. *EMBO J.* **18**, 5931–5942 (1999).

49. J. C. Silbereis *et al.*, Olig1 function is required to repress dlx1/2 and interneuron production in Mammalian brain. *Neuron*. **81**, 574–587 (2014).

50. G. J. Brewer, J. R. Torricelli, Isolation and culture of adult neurons and neurospheres. *Nat. Protoc.* **2**, 1490–1498 (2007).

51. A. Butler, P. Hoffman, P. Smibert, E. Papalexi, R. Satija, Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* **36**, 411 (2018).

52. F. Alexander Wolf, P. Angerer, F. J. Theis, SCANPY : large-scale single-cell gene expression data analysis. *Genome Biol.* **19**, 15 (2018).

53. E. Becht *et al.*, Dimensionality reduction for visualizing single-cell data using UMAP. *Nat. Biotechnol.* (2018), doi:10.1038/nbt.4314.

54. P. Held, B. Krause, R. Kruse, Dynamic Clustering in Social Networks Using Louvain and Infomap Method. *2016 Third European Network Intelligence Conference (ENIC)* (2016), , doi:10.1109/enic.2016.017.

55. G. L. Stein-O'Brien *et al.*, Decomposing Cell Identity for Transfer Learning across Cellular Measurements, Platforms, Tissues, and Species. *Cell Syst*. **8**, 395–411.e8 (2019).

56. V. Svensson, S. A. Teichmann, O. Stegle, SpatialDE: identification of spatially variable genes. *Nat. Methods*. **15**, 343 (2018).

57. A. T. L. Lun, D. J. McCarthy, J. C. Marioni, A step-by-step workflow for low-level analysis of single-cell RNA-seq data with Bioconductor. *F1000Res.* **5**, 2122 (2016).

58. X. Qiu *et al.*, Single-cell mRNA quantification and differential analysis with Census. *Nat. Methods*. **14**, 309–315 (2017).

59. A. Kataoka, T. Shimogori, Fgf8 controls regional identity in the developing thalamus. *Development*. **135**, 2873–2881 (2008).

60. C. L. Thompson *et al.*, A high-resolution spatiotemporal atlas of gene expression of the developing mouse brain. *Neuron*. **83**, 309–323 (2014).

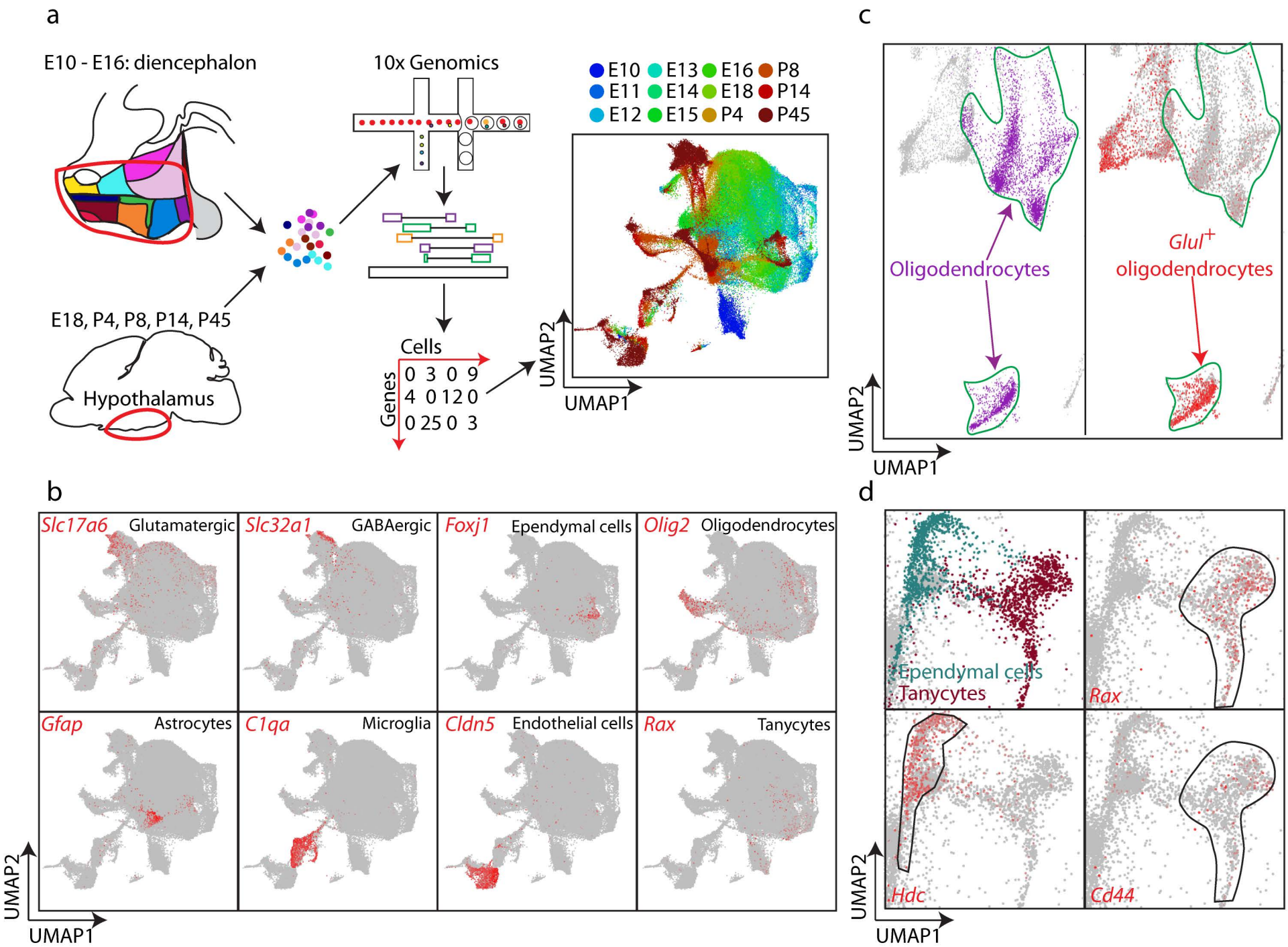

Fig1. Kim, et al.

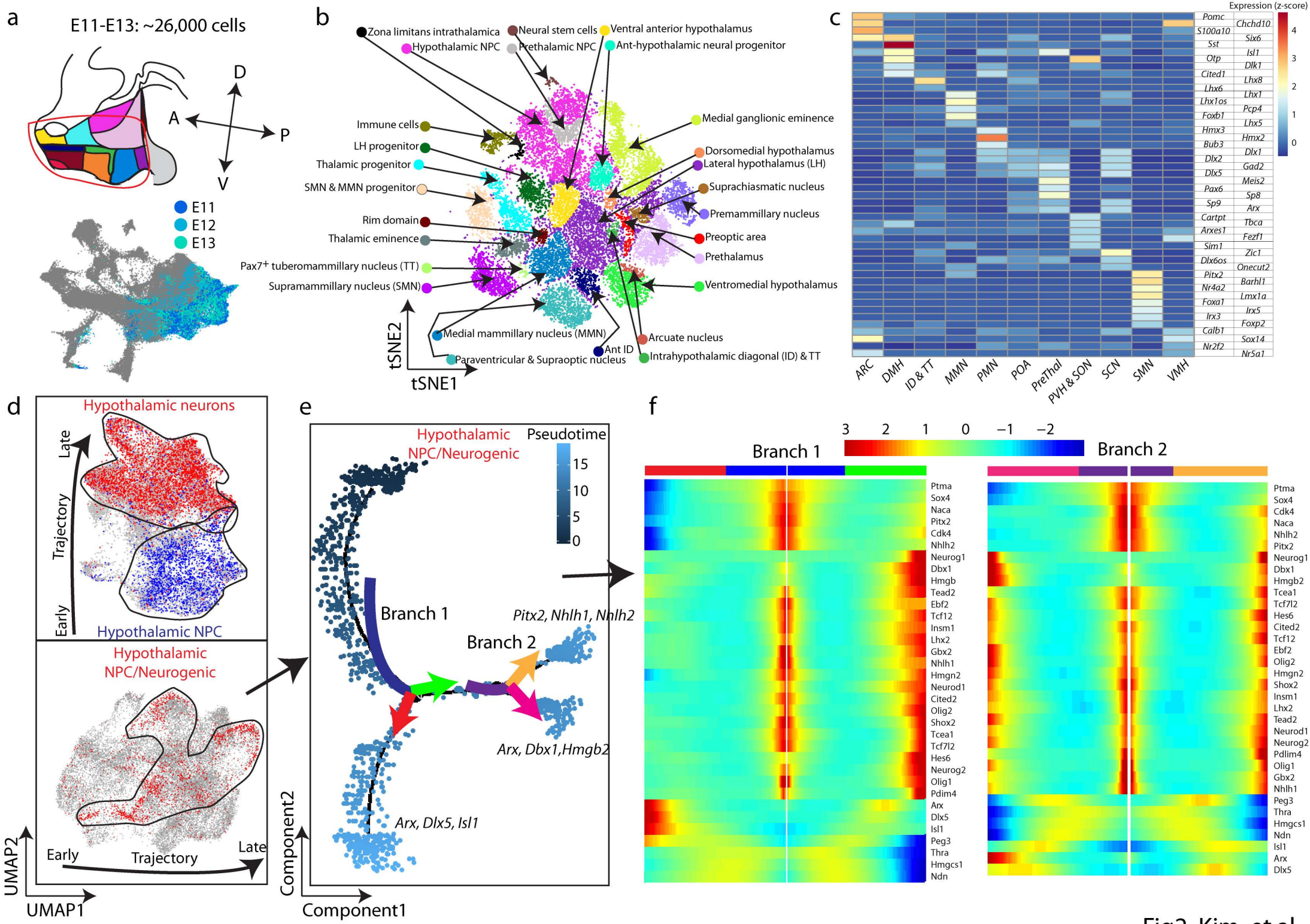

Fig2. Kim, et al.

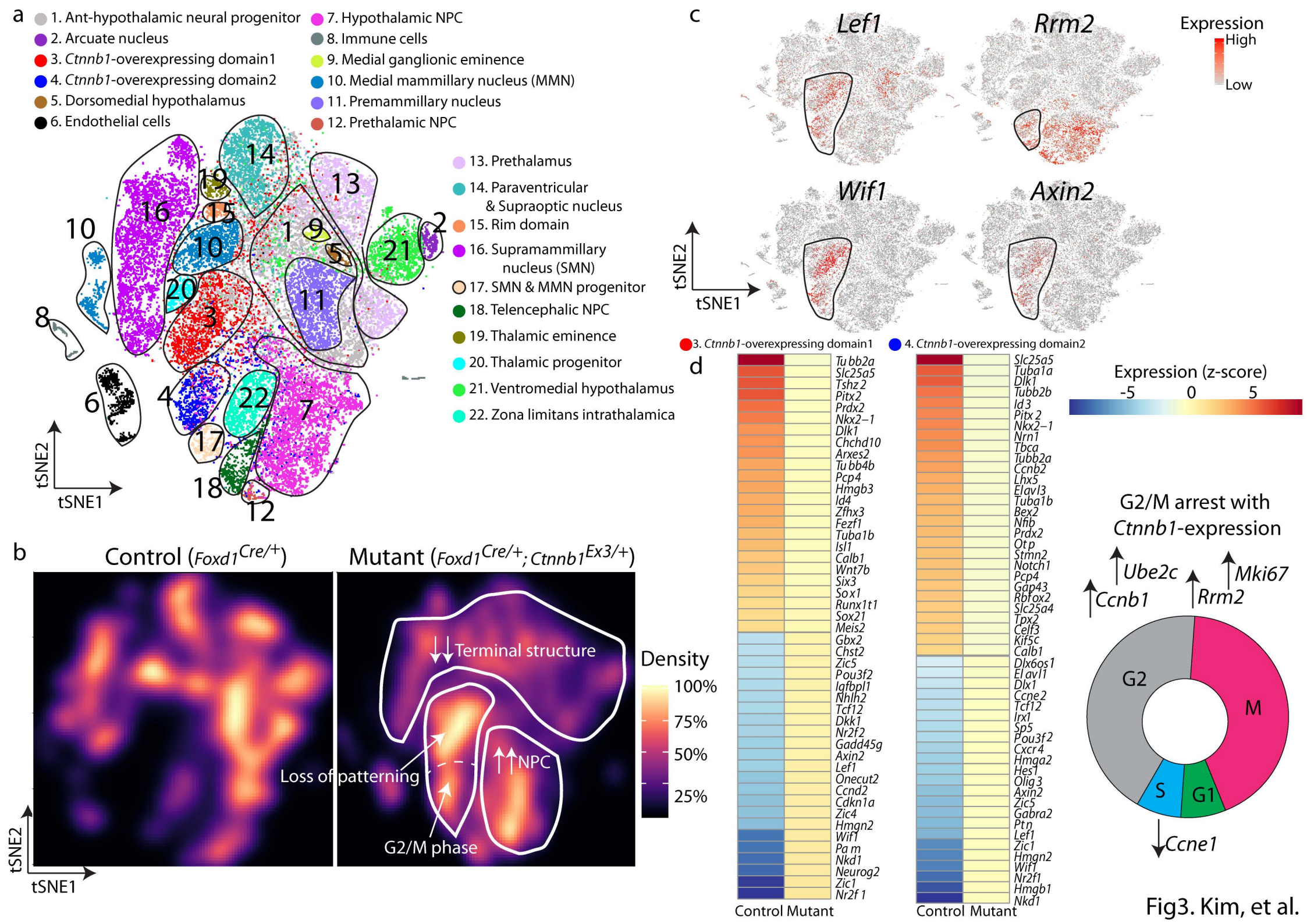

Fig3. Kim, et al.

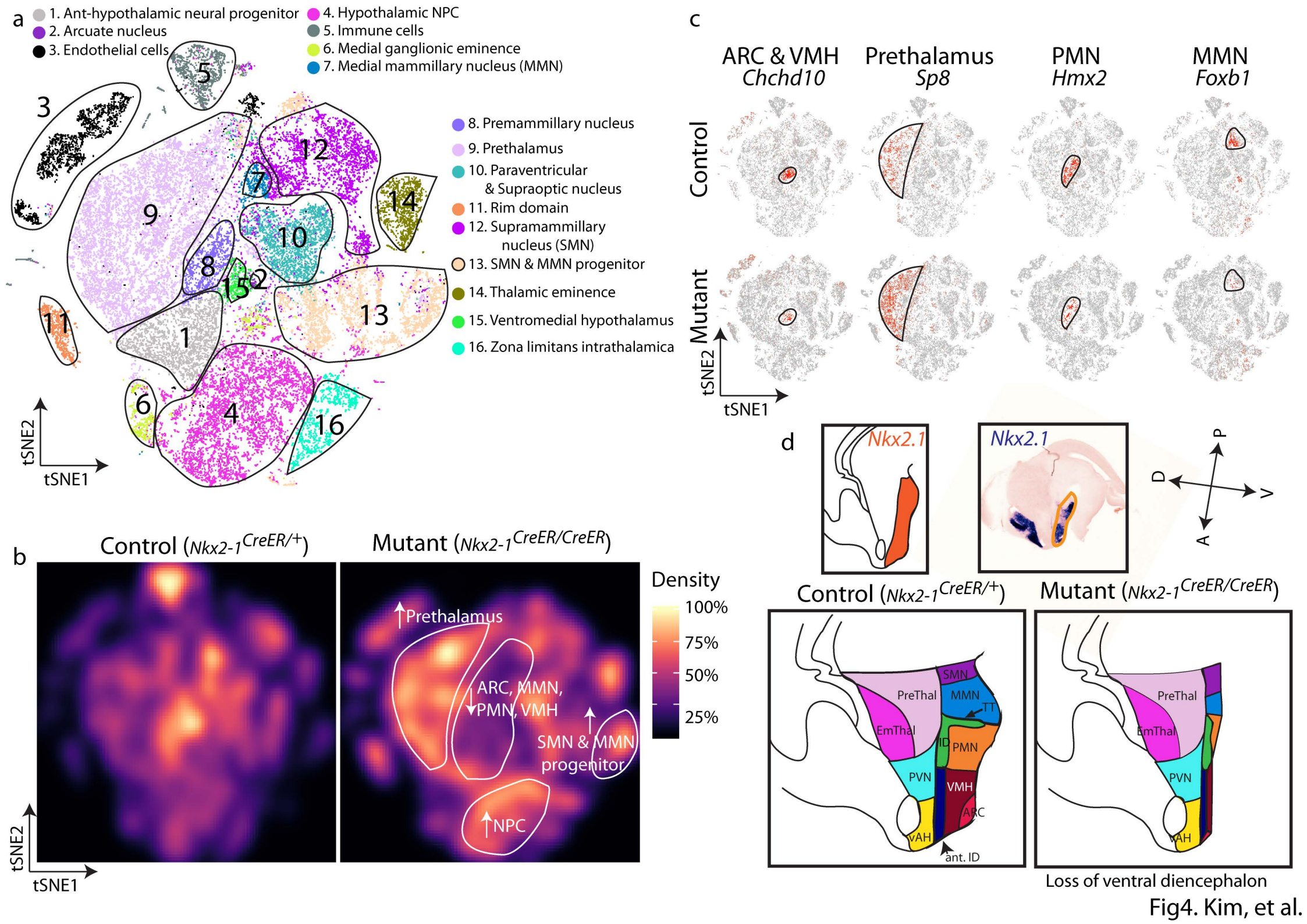

Fig4. Kim, et al.

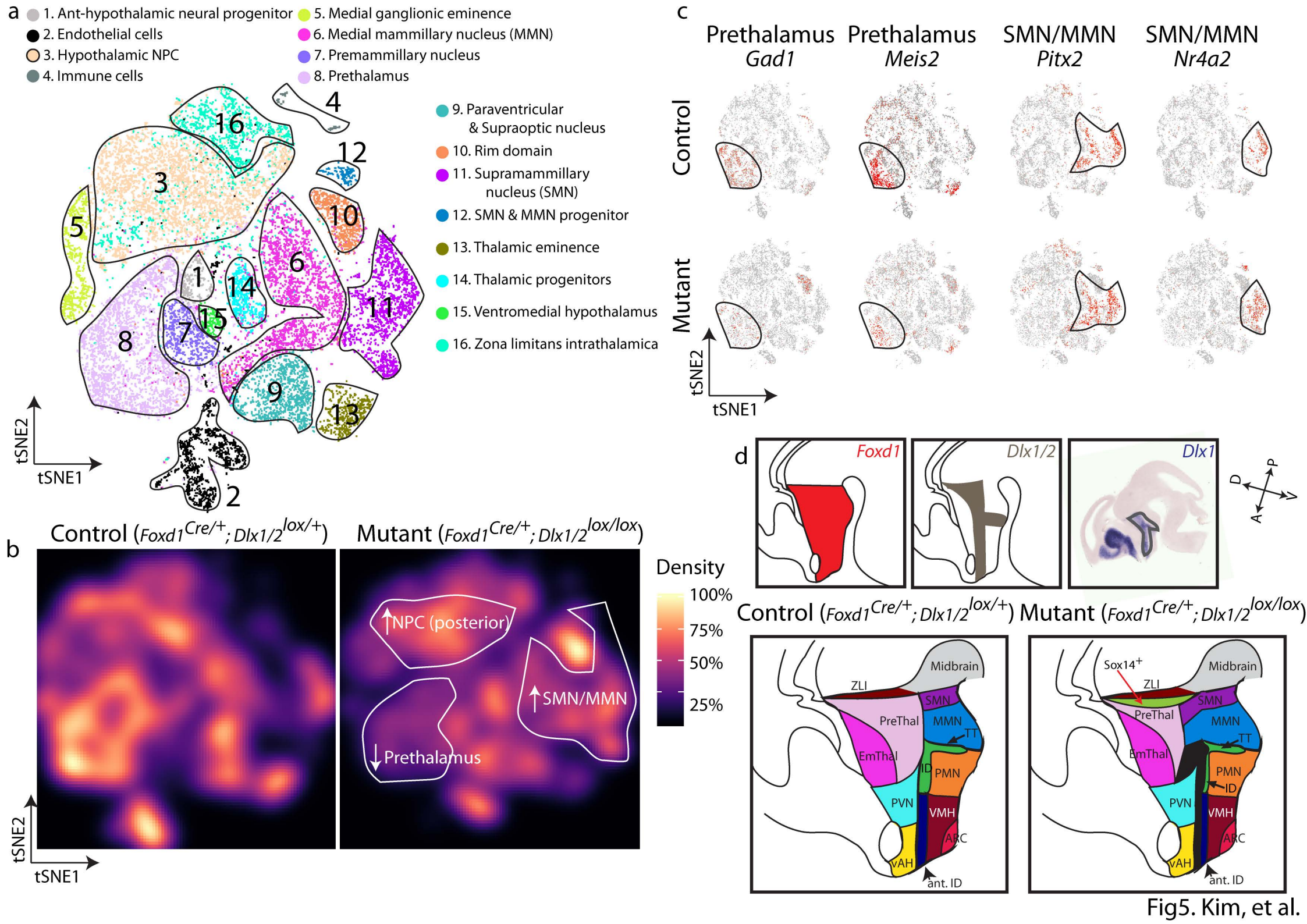

Fig5. Kim, et al.

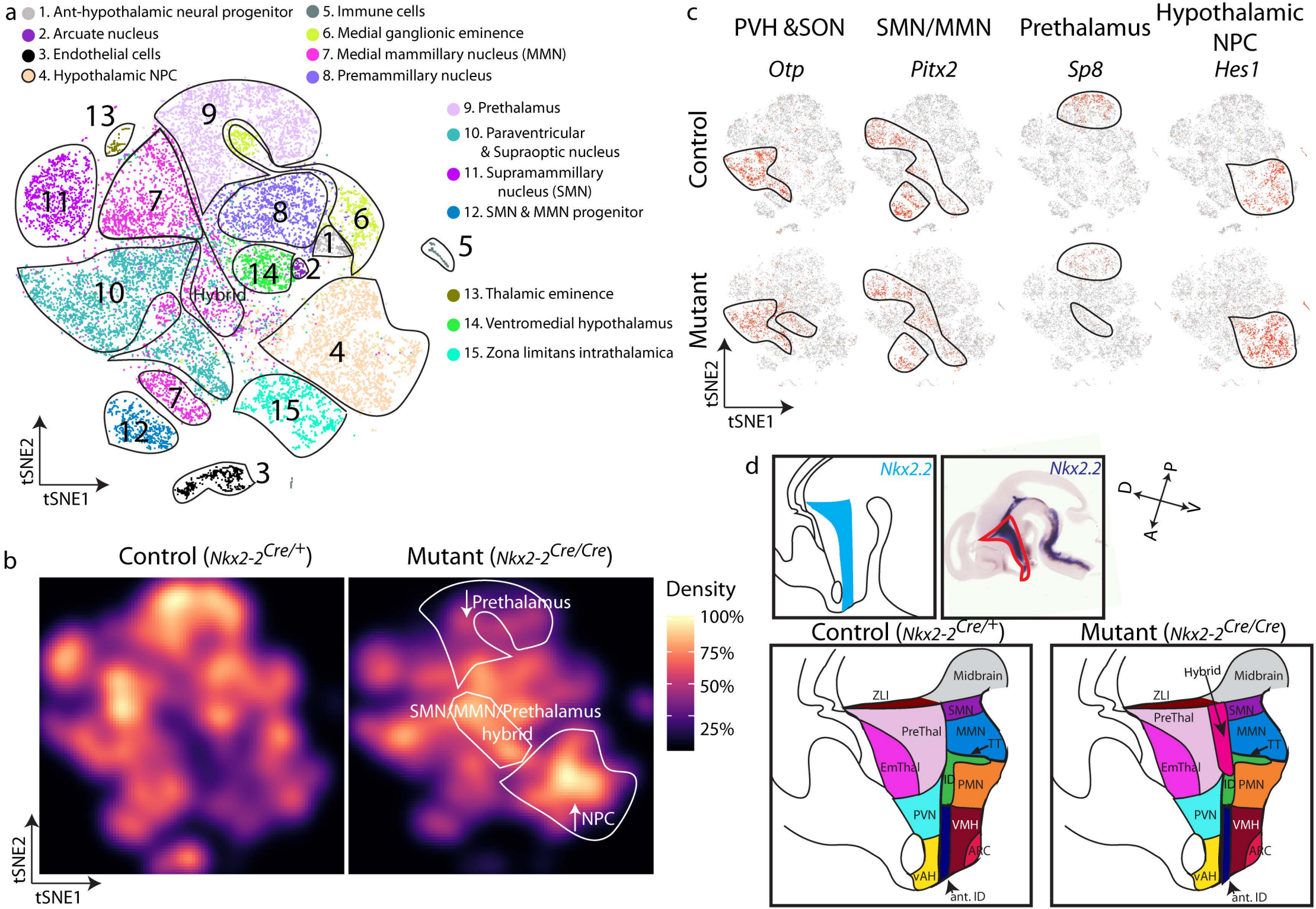

Fig6. Kim, et al.