Title: The cellular and molecular landscape of hypothalamic patterning and differentiation.

Authors: Dong Won Kim[1], Parris Whitney Washington[1*], Zoe Qianyi Wang[1*], Sonia Hao Lin[1*], Changyu Sun[1], Basma Taleb Ismail[1], Hong Wang[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 essential for survival, but the molecular mechanisms controlling hypothalamic patterning and cell fate specification are 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. By using our developmental dataset, we were able to rapidly annotate previously unidentified clusters from existing scRNA-Seq datasets collected during development, and to identify the developmental origins of major neuronal populations of the ventromedial hypothalamus. We further show that our approach can rapidly and comprehensively characterize mutants that have altered hypothalamic patterning, identifying *Nkx2.1* as a negative regulator of prethalamic identity. 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 studies 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 it possible to identify 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 remained 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,

remains contentious[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 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. In addition, 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 the development of complex organs at cellular resolution and to also 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, with a particular focus on identifying genes that control glial differentiation and function. We next focus on identifying genes that control hypothalamic regionalization and neurogenesis in the early embryo, and integrate these findings to generate a *Hy*pothalamic *D*evelopmental *D*atabase (HyDD), which identifies selective markers of each region of the developing hypothalamus and prethalamus. We next use the HyDD to rapidly annotate cell types in previously published scRNA-Seq datasets, and to infer the developmental history of specific subtypes of adult hypothalamic neurons. Finally, we demonstrate how the HyDD can be used to comprehensively analyze developmental mutants that generate complex phenotypes that would be difficult to characterize with traditional histology-based approaches, and in the process identify *Nkx2-1* as a negative regulator of prethalamic identity.

This study provides a reference atlas for future studies of hypothalamic development. It also identifies pathways by which gene regulatory networks that control cell identity can be targeted to analyze the functional role of individual hypothalamic neuronal subtypes.

Results:

Comprehensive profiling of entire course of hypothalamus development

To profile changes in gene expression across the full course of mouse hypothalamic development, we processed 12 time points 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 hypothalamic and adjacent brain regions, and major cell types at each individual age (Fig. S1c,d). Aggregation of the entire dataset with UMAP shows a clear developmental progression from neuroepithelial cells at E10, hypothalamic patterning between E11 and E13, and to the detection of major cell types in the mature hypothalamus (Fig. S2a). Roughly

similar detection of expressed genes and total mRNAs were observed at each time point (Fig. S2b,c).

Trajectories leading from neuronal progenitors to mature glutamatergic and GABAergic neurons, glia, ependymal cells and tanycytes were observed, as were separate clusters representing hematopoietic, microglial and endothelial cells (Fig. 1b). A separate cluster of oligodendrocytes was observed from P4 onwards, marked by a high and selective expression of *Glul* (Fig. 1c). Oligodendrocytes from posterior brain regions have been recently reported to selectively express high levels of *Glul*[29], suggesting that these cells may migrate into the early postnatal hypothalamus from a more posterior location. Abundant expression of hypothalamic neuropeptides including *Pomc*, *Agrp*, *Ghrh, Sst*, *Gal*, *Hcrt*, and *Pmch* were observed in the neuronal cluster (Fig. S3).

Glial cells of the hypothalamus have been shown to play critical and tissue-specific roles in regulation of osmolarity[30], circadian rhythm[31], metabolism[32] and neurogenesis[33]. To better understand the molecular mechanisms controlling the specification and differentiation of hypothalamic glia, each glial population was re-clustered and examined separately.

Cells that were identified as part of the oligodendrocyte maturation trajectory, and hence that share a similar molecular history, were re-clustered as previously described[13,34], and genes that demarcate each stage of oligodendrocyte development were identified (Fig. S4a-c, Table S1). To identify genes selectively enriched in hypothalamic oligodendrocytes, mature oligodendrocytes were directly compared to scRNA-Seq datasets from mature cortical oligodendrocytes. While *Pcsk1n* and *Cbx3* are highly enriched in hypothalamic, relative to cortical, oligodendrocytes (Fig. S4d-h), these genes are enriched in all hypothalamic glial cells, and are not specific to oligodendrocytes.

In contrast, we identified many genes that were both astrocyte-enriched relative to other glial cell types, and selectively expressed in hypothalamic, relative to cortical, astrocytes (Fig. S5a-c). These include higher expression of *Agt,* and a lower level of *Mfge8* in hypothalamic astrocytes, as previously reported[34], along with newly identified hypothalamic-enriched genes such as *Marcks* and *Marcks1* (Fig. S5a-c), which are important regulators of protein kinase C-dependent calmodulin signaling[35,36].

Analysis of the developmental trajectory connecting gliogenic progenitor cells and hypothalamic astrocytes identified transitional states between these two populations. Immature hypothalamic astrocytes co-express the mature astrocyte marker *Agt*, and *Rgcc*, a cell-cycle regulator that regulates Notch signaling[37,38] (Fig. S5d-f, Table S2). Loss of expression of genes specific to gliogenic progenitors was observed in hypothalamic astrocytes and other glial populations after the second postnatal week (Fig. S5g). Up-regulation of Notch signaling pathway components was also observed, as previously reported for human astrocyte development *in vitro*[39](Fig. S5h).

Analysis of developmental trajectories for individual hypothalamic cell types identified the age at which these cell types began to diverge in gene expression, and identified both known and candidate regulators of cell fate. This is clearly seen when comparing the development of two ventricular glial-like cell populations -- ependymal cells and tanycytes. These two classes of ventricular cells begin to diverge at E13, with

differential expression of *Foxj1* and *Rax*, markers of ependymal cells and tanycytes, first detected at this age (Fig. 1d, Fig. S6). Pseudotime analysis using BEAM analysis identifies additional transcription factors that are candidates for controlling tanycyte and ependymal cell specification and differentiation (Fig. S6a,b, Table S3). Tanycytes are themselves heterogenous, and can be subdivided into alpha and beta subtypes based on both spatial location and molecular markers[13,40]. To determine whether transcription factors enriched in differentiating tanycytes might also control tanycyte subtype specification, we analyzed previously published scRNA-Seq data obtained from mature tanycytes[14] (Fig. S6c, d), and identified multiple tanycyte subtype-specific transcription factors that are expressed during early stages of tanycyte differentiation, and hence are strong candidates for controlling tanycyte subtype specification.

Identification of genes selectively expressed in different regions of the developing hypothalamus and prethalamus scRNA-Seq

We next investigated whether we could use this dataset to faithfully distinguish hypothalamic domains that are spatially distinct in the embryo. To do this, we re-clustered data from E11, E12, and E13, which correspond to the peak period of hypothalamic neurogenesis (Fig. 2a)[41]. Using previously identified region-specific markers as a reference[2], we observed a clear segregation of spatially-distinct neuronal precursors and progenitors (Fig. 2b, S7). We were able to readily distinguish hypothalamic and adjacent cell populations including the prethalamus, discrete clusters for telencephalic structures such as preoptic area and medial ganglionic eminence, thalamic eminence, rim domain, and main body of the sensory thalamus, as well as the zona limitans intrathalamica (ZLI) at all three developmental ages (Fig. 2b, S7-S8).

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), extrahypothalamic diagonal (ID) and tuberomammillary terminal (TT), ventromedial hypothalamus (VMH), arcuate nucleus (ARC), premammillary hypothalamus (PMN), mammillary nucleus (MMN), and supramammillary nucleus (SMN) (Fig. 2c,d, S7). 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. 2d, Table S4).

Multiple known and previously undescribed molecular markers, including many transcription factors, were identified for each of these regions (Fig. 2c, Table S4). While some of these markers are shared among multiple regions in the hypothalamus and other forebrain regions, others are highly specific and non-overlapping. We identified clear separation between mitotic neural progenitors and postmitotic neural precursors (Fig. 2d, S7).

This analysis was able to efficiently identify gene expression patterns that were 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 expression patterns, but which nevertheless may 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[42] (Fig. S8). Using this method, we identified patterns that not only matched key spatial

185 subdivisions of the hypothalamus and prethalamus, but also patterns that labeled
186 discrete subsets of hypothalamic progenitors (Table S5). These include hypothalamic
187 neural precursor cells (NPC) that likely correspond to radial glia (*Fabp7*, *Slc1a3*), as
188 well as neurogenic progenitors (*Pitx2*, *Nhlh1*, *Nhlh2*). Most strikingly, we observed
189 multiple patterns that selectively label neurogenic progenitors that are located along the
190 borders of the hypothalamus and prethalamus with both the telencephalon and ZLI.
191 Genes that drive this pattern include *Neurog2*, *Lhx5*, and *Nhlh2* (Table S5). Although
192 these expression patterns have been previously reported[2], the fate of these border cells
193 is unknown.
194 Due to the high complexity of the hypothalamic clusters observed in both two-
195 and three-dimensional analysis, it is difficult to comprehensively visualize region-specific
196 differences in gene expression. To improve visualization of these data, we generated a
197 heatmap for major pattern marker genes that corresponds to the two-dimensional
198 sagittal plane, capturing the main spatial subdivisions of the developing hypothalamus
199 and adjacent brain regions (Fig. S9).
200 This analysis also identified clusters that correspond to three hypothalamic
201 regions that had not been described in previous work[2], including two populations of
202 excitatory neurons. The first of these regions is found in the dorsomedial
203 hypothalamus, and is marked by expression of *Sst*, *Cited1*, *Otp* and *Six6* (Fig. 2c). The
204 second region is found in the TT/PMN region, and expresses *Pax7* (Table. S4). The
205 third region is found in the lateral hypothalamus (LH), and consists of a diverse
206 collection of subtypes of neuronal precursors. This LH cluster consists primarily of
207 glutamatergic neurons, with a small subpopulation of GABAergic neurons (Fig. S10,
208 Table S6). The glutamatergic population includes a discrete subcluster of *Lhx9*-positive
209 neurons, which marks precursors of hypocretin neurons[2,43,44]. Cells within this LH
210 cluster express multiple transcription factors that are also selectively expressed in other
211 hypothalamic regions, including the VMH, PMN, MMN and ID.
212 Clustering of previously characterized spatial domains also identified discrete
213 subclusters that express common sets of genes. This is clearly seen in the PVN/SON
214 cluster (Fig. S11, Table S6). Selective expression of *Onecut2*, *Cartpt* and *Zic1*
215 characterizes a ventrolateral domain that, based on its position, likely corresponds to
216 the developing SON (Fig. S11). This same approach can be readily applied to other
217 forebrain regions. We have previously identified molecular markers that both identify
218 discrete spatial domains within the prethalamus, which gives rise to structures such as
219 the thalamic reticular nucleus and ventral lateral geniculate nucleus[2,45,46], and
220 investigated whether these regions could be identified using scRNA-Seq data.
221 Sub-clustering of prethalamic cells allowed us to detect these and other spatial
222 subdivisions within the prethalamus. We observed partially overlapping domains of
223 expression of the transcription factors *Sp8* and *Sp9* (Fig. S12, Table S6), which play
224 critical roles in the development of telencephalic interneurons[47]. *In situ* hybridization
225 analysis revealed enriched expression of *Sp8* and *Lhx1* in anterior prethalamus and ID,
226 while *Sp9* and *Prox1* were enriched in posterior prethalamus (Fig. S12). *Zic1* and *Ebf*1
227 also marked distinct, but partially overlapping spatial domains in prethalamus (Fig. S12).
228 Sub-clustering of the VMH allowed us to detect two distinct clusters, which
229 corresponded to separate anterior and posterior domains of gene expression (Fig. S13,
230 Table S6). A clear distinction between these anterior and posterior domains was

231 detected until E16, both spatially and at the molecular level (Fig. S13). These two
232 clusters had begun to spatially intermingle, yet the molecular distinction still remained,
233 possibly reflecting local tangential cell migration within the VMH.
234 By combining our analysis of both the molecular markers of differentiation of
235 major hypothalamic cell types and the selective markers of the different spatial domains
236 of the developing hypothalamus and prethalamus, we have compiled a reference set of
237 molecular markers that will be useful for further functional studies. We have
238 designated this integrated and annotated scRNA-Seq dataset as a HyDD, or the
239 *Hy*pothalamus *D*evelopmental *D*atabase.
240
241 **HyDD can rapidly annotate existing hypothalamus scRNA-Seq dataset and**
242 **identifies developmental origins of VMH neurons**
243 To demonstrate the broad usefulness of the HyDD, we first annotated a
244 previously published scRNA-Seq dataset obtained through selective dissection of
245 *Pomc-EGFP*-expressing cells from E15.5 hypothalamus using regional and cell type-
246 specific markers from the HyDD[48]. In this study, while one cluster (cluster 0) was
247 previously identified as the developing ARC, the remaining clusters were not annotated
248 owing to the lack of well defined regional and cell type-specific markers to resolve
249 spatial location of these clusters. Using markers obtained from the HyDD to train the
250 dataset, we were able to annotate all but two clusters, representing cells from multiple
251 hypothalamic regions, including VMH, PMH, anterior ID, DMH, SCN, and ARC. Some
252 clusters were composed of cells from multiple hypothalamic regions, which may explain
253 some of the previous difficulties in annotating these cells (Fig. 3a,b). Two unannotated
254 clusters appear to reflect contamination from the habenula and pituitary that occurred
255 during dissection (Fig. S14). A subset of the neurons in the ARC cluster share
256 molecular markers of neural precursors in the PMN and DMH, implying that these cells
257 may have migrated to the ARC from these regions (Fig. S14).
258 We also identified a cluster that closely resembled hypothalamic NPC (Fig. S15),
259 but which also co-expressed astrocyte-, ependymal and/or tanycyte-specific marker
260 genes. Gene sets enriched in this cluster were then projected into the entire
261 hypothalamus scRNA-Seq dataset (E10-P45), and glial populations including immature
262 glial cells were enriched with these gene sets. This same gene expression pattern was
263 found to be enriched in a subset of hypothalamic NPC that were detected from E11
264 onwards, and which may represent NPC that are competent to generate glia (Fig. S15).
265 Many of these same genes are also expressed in the late-stage retinal progenitor cells,
266 from the age at which they become competent to give rise to tanycyte-like Mũller glial
267 cells[28].
268 Since HyDD contains a nearly uninterrupted temporal profile of changes in gene
269 expression during the process of cell specification and differentiation, HyDD can also be
270 used to infer the developmental origins of fully mature hypothalamic neurons. However,
271 identifying the precise spatial location of individual cell types from hypothalamus
272 scRNA-Seq data based on specific molecular markers alone is bioinformatically
273 challenging, due to the extreme tissue complexity. This is the case even when scRNA-
274 Seq data has been generated with micro-dissected or flow-sorted cells from pre-defined
275 hypothalamic regions. Most informative region-specific markers are strongly expressed
276 early in development, but are either not expressed or show substantially different

expression levels at later developmental ages[49]. Postmitotic hypothalamic neural precursors also undergo a considerable amount of tangential migration and dispersion, making it even harder to directly identify gene regulatory networks that control the specification of individual hypothalamic cell types[50].

To identify the developmental origin of individual hypothalamic cell types, it is critical that overlapping sets of markers be identified that selectively label each stage of cell differentiation, in a manner analogous to molecular stepping stones, so that the developmental history of each cell type can be reconstructed. As a proof of principle for this approach, we identified gene sets that identify VMH cells at early stages of hypothalamic development (Fig. 3C), when region-specific molecular markers are robustly expressed. Gene sets specific to discrete spatial domains were then used to train the following developmental age to find VMH cells and new VMH-enriched genes were identified. This process was repeated for each successive developmental age. These VMH-enriched genes have varying levels of expression and specificity across the full course of the hypothalamus development (Fig. S16).

We next used the HyDD to identify the developmental origin of major VMH neuronal subtypes. Recent scRNA-Seq of the adult VMH identified multiple clusters of both core glutamatergic VMH neurons and of GABAergic neurons surrounding the core VMH (VMH-out)[51]. We sought to identify the developmental origins of both classes of VMH neurons. We first found that GABAergic neurons of VMH-out originated from four distinct regions of the developing hypothalamus - ARC, DMH, Ant ID and PMN (Fig. 3d) -- with each VMH-out GABAergic cluster having a distinct developmental origin based on specific expression of regional markers. We likewise observed that different subsets of core glutamatergic VMH neurons arise from distinct anterior or posterior domains of the embryonic VMH (Fig. S13). Some of these clusters remain restricted to anterior or posterior regions of the adult VMH, as noted in the original study[51](Fig. 3d, Fig. S17a,b). However, the majority of VMH neuronal subtypes originate from both anterior and posterior domains of the developing VMH (Fig. S17), and are distributed widely along the anterior-posterior axis of the adult VMH[51]. VMH neuronal subtypes may thus be two distinct developmental steps: an initial stage in which anterior and posterior identity is specified between E11 and E13, and a later stage that coincides with the initiation of local tangential migration that occurs from E16 onwards.

HyDD allow rapid and comprehensive analysis of complex mutant phenotypes that alter hypothalamic and prethalamic patterning

HyDD provides both a high-resolution molecular atlas of the developing hypothalamus and prethalamus, and a useful resource to understand the developmental origin of adult hypothalamic neurons. We next sought to determine if HyDD could also be used to rapidly and comprehensively characterize mutants that regulate early stages of hypothalamic development and organization. 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[20]. The same analysis was also with *Foxd1*$^{CreGFP/+}$ 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

323 hypothalamic markers, show the loss of most regional markers in the hypothalamus and
324 prethalamus[20].
325 ScRNA-Seq analysis of control and mutant animals at E12.5 reveals several
326 mutant-specific cell clusters (Fig. S18a). Using the HyDD to annotate both control and
327 mutant data, we identified changes in gene expression and cell composition that match
328 previously reported findings (Fig. 4a), where we observed a substantial increase in
329 undifferentiated NPC, along with a corresponding reduction in the number of cells
330 expressing markers of hypothalamic and prethalamic neuronal precursors (Fig. 4b,
331 S18). In particular, strong loss of markers shared by both hypothalamus and
332 prethalamus, such as *Arx*, *Isl1* and *Gad1/2* (Fig. S19, Table S7) was observed. We
333 also identified two cell clusters that are found exclusively in mutant mice, both of which
334 express NPC markers, and also highly express both *Lef1* and negative regulators of
335 canonical Wnt signaling such as *Dkk1*, *Wif1* and *Axin2* (Fig. S18e). One of these
336 clusters is strongly enriched for G2/M phase markers such *Ube2c*, *Rrm2,* and *Ccnb1*
337 (Fig. 4b). Flow cytometry data also demonstrated a substantially higher fraction of
338 NPCs in G2/M phase in mutant mice (Fig. S18f), as has been previously reported in
339 non-neuronal cells that show high levels of canonical Wnt signaling[52]. This finding
340 explains the previous observation that, although a massive increase in the number of
341 NPC cells is seen in these mutants, only a modest increase is observed in EdU labeling,
342 which labels S-phase NPC[20]. This demonstrates the power of using scRNA-Seq in
343 conjunction with the HyDD to analyze developmental phenotypes, in a manner that is
344 far more rapid and comprehensive than conventional histological techniques.
345 We next used this same approach to characterize E12.5 *Nkx2-1*$^{CreER/CreER}$ knock-
346 in mice, which are homozygous for a null allele in the homeodomain transcription factor
347 *Nkx2-1*[53]. *Nkx2-1* is broadly and selectively expressed in ventral hypothalamic
348 progenitors, as well as in progenitors that give rise to telencephalic interneurons[54,55].
349 Loss of function of *Nkx2-1* leads to a substantial reduction in ventral hypothalamic
350 structures by E18[56], but a detailed molecular characterization of these mutants has not
351 been conducted.
352 Analysis of *Nkx2-1*$^{CreER/CreER}$ mutants and heterozygous littermate controls
353 revealed changes in cluster densities in the mutant (Fig. S20). We observed a broad
354 loss of markers specific to *Nkx2-1* positive ventral hypothalamic structures such as
355 ARC, VMH, PMN, and MMN, but not the SMN (Fig. 4c,d, Fig. S21, Table S8, S9), with
356 both the relative expression levels and the number of cells expressing these markers
357 reduced. Both the width of the hypothalamic ventricular zone and the levels of EdU
358 incorporation were reduced in *Nkx2-1*$^{CreER/CreER}$ mice (Fig. S22). An increase in the
359 fraction of cells expressing prethalamic markers was detected (Fig. 4d, Fig. S23), and
360 increased Cre expression was also observed in these mice.
361 In contrast to controls, prethalamic cells in mutant mice expressed *Cre*, implying
362 that ventral hypothalamic cells that normally express *Nkx2.1* may have acquired
363 prethalamic identity (Fig. S22). To investigate this further, RNAscope probes against
364 *Sp9*, *Meis2* and *Cre* were used to visualize the location of these *Cre*-positive
365 prethalamic cells, and substantial co-localization of prethalamic markers and *Cre*
366 expression was observed in the region normally occupied the by the ventral
367 hypothalamus in controls (Fig. S22). This implies that *Nkx2-1* not only maintains the
368 identity of ventral hypothalamic progenitors but also actively represses expression of

molecular markers of prethalamic identity. *In situ* hybridization confirmed that there was an increase in the absolute size of the prethalamus and its proportion in the diencephalon (Fig. S23). An increase in the number of cells expressing markers of NPC in the SMN and MMN was also seen, while *Nkx2-1* negative hypothalamic regions such as the PVN/SON are unaffected (Fig. 4d, S24).

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 accurately delineates spatial subdivisions present in the early stages of development of both the hypothalamus and the adjacent prethalamus. It also identifies many previously uncharacterized transcription factors and other genes that are excellent candidates for controlling regional patterning and specification of individual hypothalamic cell types. Combining functional analysis of these genes with the new selective markers of hypothalamic regions and immature hypothalamic cell types identified in this study has the promise to greatly expand our knowledge of hypothalamic development and organization.

The integrated dataset presented here provides three specific features that are critical for studying the formation and function of the hypothalamus. First, it makes it straightforward to unambiguously annotate major cell types at all stages of hypothalamic development. Second, it makes it possible in many cases to infer the developmental histories of hypothalamic cells in both the developing and mature hypothalamus. Third, it allows rapid and accurate phenotyping of mutants that show broad effects on hypothalamic patterning, neurogenesis and differentiation, with which we were able to validate our findings using 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. The HyDD dataset allows these analyses to be conducted far more rapidly, efficiently, and comprehensively.

Our scRNA-Seq characterization of *Nkx2-1*-deficient mice identifies an unexpected developmental connection between the hypothalamus and prethalamus, where *Nkx2-1* can potentially act as both a positive regulator of ventral hypothalamic identity while simultaneously repressing prethalamic identity. This result is not predicted by the current prosomeric model for forebrain organization[57,58], and raises questions about the early development and patterning of these structures. Previous models of hypothalamic development and organization were constructed using very sparse datasets - typically single color in situ hybridization of a limited number of genes at a small number of time points. The much richer datasets provided by scRNA-Seq, and interpreted using the HyDD data, offer a far more powerful resource for constructing these models..

References:

1. Bedont, J. L., Newman, E. A. & Blackshaw, S. Patterning, specification, and differentiation in the developing hypothalamus. *Wiley Interdiscip. Rev. Dev. Biol.* **4**, 445–468 (2015).
2. Shimogori, T. *et al.* A genomic atlas of mouse hypothalamic development. *Nat. Neurosci.* **13**, 767–775 (2010).
3. Xie, Y. & Dorsky, R. I. Development of the hypothalamus: conservation, modification and innovation. *Development* **144**, 1588–1599 (2017).
4. Kent, M. A. & Peters, R. H. Effects of ventromedial hypothalamic lesions on hunger-motivated behavior in rats. *J. Comp. Physiol. Psychol.* **83**, 92–97 (1973).
5. Kruk, M. R. *et al.* Discriminant analysis of the localization of aggression-inducing electrode placements in the hypothalamus of male rats. *Brain Research* vol. 260 61–79 (1983).
6. Lammers, J. H., Kruk, M. R., Meelis, W. & van der Poel, A. M. Hypothalamic substrates for brain stimulation-induced attack, teeth-chattering and social grooming in the rat. *Brain Res.* **449**, 311–327 (1988).
7. Hervey, G. R. The effects of lesions in the hypothalamus in parabiotic rats. *J. Physiol.* **145**, 336–352 (1959).
8. Bedont, J. L. *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. Kohl, J. *et al.* Functional circuit architecture underlying parental behaviour. *Nature* **556**, 326–331 (2018).
10. Yang, T. *et al.* Social Control of Hypothalamus-Mediated Male Aggression. *Neuron* **95**, 955–970.e4 (2017).
11. Romanov, R. A. *et al.* Molecular interrogation of hypothalamic organization reveals distinct dopamine neuronal subtypes. *Nat. Neurosci.* **20**, 176–188 (2017).
12. Mickelsen, L. E. *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. Chen, R., Wu, X., Jiang, L. & Zhang, Y. Single-Cell RNA-Seq Reveals Hypothalamic Cell Diversity. *Cell Reports* vol. 18 3227–3241 (2017).
14. Campbell, J. N. *et al.* A molecular census of arcuate hypothalamus and median eminence cell types. *Nat. Neurosci.* **20**, 484–496 (2017).
15. Swanson, L. W. *Brain Maps: Structure of the Rat Brain*. (Elsevier Publishing Company, 1992).
16. Kuhlenbeck, H. *The Central Nervous System of Vertebrates: pt. 1. Structural elements: biology of nervous tissue. pt. 2. Overall morphologic pattern*. (1967).
17. Rubenstein, J. L., Martinez, S., Shimamura, K. & Puelles, L. The embryonic vertebrate forebrain: the prosomeric model. *Science* **266**, 578–580 (1994).

455 18. Visel, A., Thaller, C. & Eichele, G. GenePaint.org: an atlas of gene expression
456 patterns in the mouse embryo. *Nucleic Acids Res.* **32**, D552 (2004).
457 19. Lein, E. S. *et al.* Genome-wide atlas of gene expression in the adult mouse brain.
458 *Nature* **445**, 168–176 (2007).
459 20. Newman, E. A., Wu, D., Taketo, M. M., Zhang, J. & Blackshaw, S. Canonical Wnt
460 signaling regulates patterning, differentiation and nucleogenesis in mouse
461 hypothalamus and prethalamus. *Dev. Biol.* **442**, 236–248 (2018).
462 21. Newman, E. A. *et al.* Foxd1 is required for terminal differentiation of anterior
463 hypothalamic neuronal subtypes. *Dev. Biol.* **439**, 102–111 (2018).
464 22. Kurrasch, D. M. *et al.* The neonatal ventromedial hypothalamus transcriptome
465 reveals novel markers with spatially distinct patterning. *J. Neurosci.* **27**, 13624–
466 13634 (2007).
467 23. Lee, B. *et al.* Dlx1/2 and Otp coordinate the production of hypothalamic GHRH- and
468 AgRP-neurons. *Nat. Commun.* **9**, 2026 (2018).
469 24. Sokolowski, K. *et al.* Specification of select hypothalamic circuits and innate
470 behaviors by the embryonic patterning gene dbx1. *Neuron* **86**, 403–416 (2015).
471 25. Liu, K. *et al.* Lhx6-positive GABA-releasing neurons of the zona incerta promote
472 sleep. *Nature* **548**, 582–587 (2017).
473 26. Bedont, J. L. *et al.* Lhx1 controls terminal differentiation and circadian function of
474 the suprachiasmatic nucleus. *Cell Rep.* **7**, 609–622 (2014).
475 27. Stuart, T. & Satija, R. Integrative single-cell analysis. *Nature Reviews Genetics* vol.
476 20 257–272 (2019).
477 28. Clark, B. S. *et al.* Single-Cell RNA-Seq Analysis of Retinal Development Identifies
478 NFI Factors as Regulating Mitotic Exit and Late-Born Cell Specification. *Neuron*
479 (2019) doi:10.1016/j.neuron.2019.04.010.
480 29. Xin, W. *et al.* Oligodendrocytes Support Neuronal Glutamatergic Transmission via
481 Expression of Glutamine Synthetase. *Cell Rep.* **27**, 2262–2271.e5 (2019).
482 30. Choe, K. Y., Olson, J. E. & Bourque, C. W. Taurine release by astrocytes
483 modulates osmosensitive glycine receptor tone and excitability in the adult
484 supraoptic nucleus. *J. Neurosci.* **32**, 12518–12527 (2012).
485 31. Tso, C. F. *et al.* Astrocytes Regulate Daily Rhythms in the Suprachiasmatic Nucleus
486 and Behavior. *Curr. Biol.* **27**, 1055–1061 (2017).
487 32. Chowen, J. A. *et al.* The role of astrocytes in the hypothalamic response and
488 adaptation to metabolic signals. *Prog. Neurobiol.* **144**, 68–87 (2016).
489 33. Lee, D. A. *et al.* Tanycytes of the hypothalamic median eminence form a diet-
490 responsive neurogenic niche. *Nature Neuroscience* vol. 15 700–702 (2012).
491 34. Zeisel, A. *et al.* Molecular Architecture of the Mouse Nervous System. *Cell* **174**,
492 999–1014.e22 (2018).
493 35. Gallant, C., You, J. Y., Sasaki, Y., Grabarek, Z. & Morgan, K. G. MARCKS is a
494 major PKC-dependent regulator of calmodulin targeting in smooth muscle. *J. Cell*

Sci. **118**, 3595–3605 (2005).
36. El Amri, M., Fitzgerald, U. & Schlosser, G. MARCKS and MARCKS-like proteins in development and regeneration. *J. Biomed. Sci.* **25**, 43 (2018).
37. Counts, S. E. & Mufson, E. J. Regulator of Cell Cycle (RGCC) Expression During the Progression of Alzheimer's Disease. *Cell Transplant.* **26**, 693–702 (2017).
38. Carrieri, F. A. *et al.* CDK1 and CDK2 regulate NICD1 turnover and the periodicity of the segmentation clock. *EMBO Rep.* **20**, e46436 (2019).
39. Malik, N. *et al.* Comparison of the gene expression profiles of human fetal cortical astrocytes with pluripotent stem cell derived neural stem cells identifies human astrocyte markers and signaling pathways and transcription factors active in human astrocytes. *PLoS One* **9**, e96139 (2014).
40. Goodman, T. & Hajihosseini, M. K. Hypothalamic tanycytes—masters and servants of metabolic, neuroendocrine, and neurogenic functions. *Frontiers in Neuroscience* vol. 9 (2015).
41. Shimada, M. & Nakamura, T. Time of neuron origin in mouse hypothalamic nuclei. *Exp. Neurol.* **41**, 163–173 (1973).
42. Stein-O'Brien, G. L. *et al.* Decomposing cell identity for transfer learning across cellular measurements, platforms, tissues, and species. *bioRxiv* 395004 (2018) doi:10.1101/395004.
43. Liu, J. *et al.* Evolutionarily conserved regulation of hypocretin neuron specification by Lhx9. *Development* **142**, 1113–1124 (2015).
44. Dalal, J. *et al.* Translational profiling of hypocretin neurons identifies candidate molecules for sleep regulation. *Genes Dev.* **27**, 565–578 (2013).
45. Vue, T. Y. *et al.* Characterization of progenitor domains in the developing mouse thalamus. *J. Comp. Neurol.* **505**, 73–91 (2007).
46. Jeong, Y. *et al.* Spatial and temporal requirements for sonic hedgehog in the regulation of thalamic interneuron identity. *Development* **138**, 531–541 (2011).
47. Li, J. *et al.* Transcription Factors Sp8 and Sp9 Coordinately Regulate Olfactory Bulb Interneuron Development. *Cereb. Cortex* **28**, 3278–3294 (2018).
48. Huisman, C. *et al.* Single cell transcriptome analysis of developing arcuate nucleus neurons uncovers their key developmental regulators. *Nat. Commun.* **10**, 3696 (2019).
49. Shimogori, T. *et al.* A genomic atlas of mouse hypothalamic development. *Nat. Neurosci.* **13**, 767–775 (2010).
50. Arnold-Aldea, S. A. & Cepko, C. L. Dispersion Patterns of Clonally Related Cells during Development of the Hypothalamus. *Developmental Biology* vol. 173 148–161 (1996).
51. Kim, D.-W. *et al.* Multimodal Analysis of Cell Types in a Hypothalamic Node Controlling Social Behavior. *Cell* vol. 179 713–728.e17 (2019).
52. Olmeda, D., Castel, S., Vilaró, S. & Cano, A. β-Catenin Regulation during the Cell

535 Cycle: Implications in G2/M and Apoptosis. *Molecular Biology of the Cell* vol. 14
536 2844–2860 (2003).
537 53. Taniguchi, H. *et al.* A resource of Cre driver lines for genetic targeting of GABAergic
538 neurons in cerebral cortex. *Neuron* **71**, 995–1013 (2011).
539 54. Du, T., Xu, Q., Ocbina, P. J. & Anderson, S. A. NKX2.1 specifies cortical
540 interneuron fate by activating Lhx6. *Development* **135**, 1559–1567 (2008).
541 55. Elias, L. A. B., Potter, G. B. & Kriegstein, A. R. A time and a place for nkx2-1 in
542 interneuron specification and migration. *Neuron* vol. 59 679–682 (2008).
543 56. Kimura, S. *et al.* The T/ebp null mouse: thyroid-specific enhancer-binding protein is
544 essential for the organogenesis of the thyroid, lung, ventral forebrain, and pituitary.
545 *Genes Dev.* **10**, 60–69 (1996).
546 57. Puelles, L. Forebrain Development: Prosomere Model. *Encyclopedia of*
547 *Neuroscience* 315–319 (2009) doi:10.1016/b978-008045046-9.01076-7.
548 58. Ferran, J. L., Puelles, L. & Rubenstein, J. L. R. Molecular codes defining
549 rostrocaudal domains in the embryonic mouse hypothalamus. *Frontiers in*
550 *Neuroanatomy* vol. 9 (2015).

551

Contribution: DWK, SB designed experiments. DWK, PWW, ZQW, BTI, SL, LJ, HW performed experiments. DWK, PWW, ZQW, SL, CS analyzed data. All authors contributed to writing the paper

Acknowledgements:
This work was supported by a grant from the NIH (DK108230) to S.B. We thank Transcriptomics and Deep Sequencing Core at Johns Hopkins for sequencing all scRNA-Seq libraries, Ross Flow Cytometry Core (Johns Hopkins) for FACS analysis, and Microscope facility (Johns Hopkins MICFAC, supported by the award number S10OD018118). 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, GSE132355. Data can be viewed at https://proteinpaint.stjude.org/F/mm10/example.scrna.html.

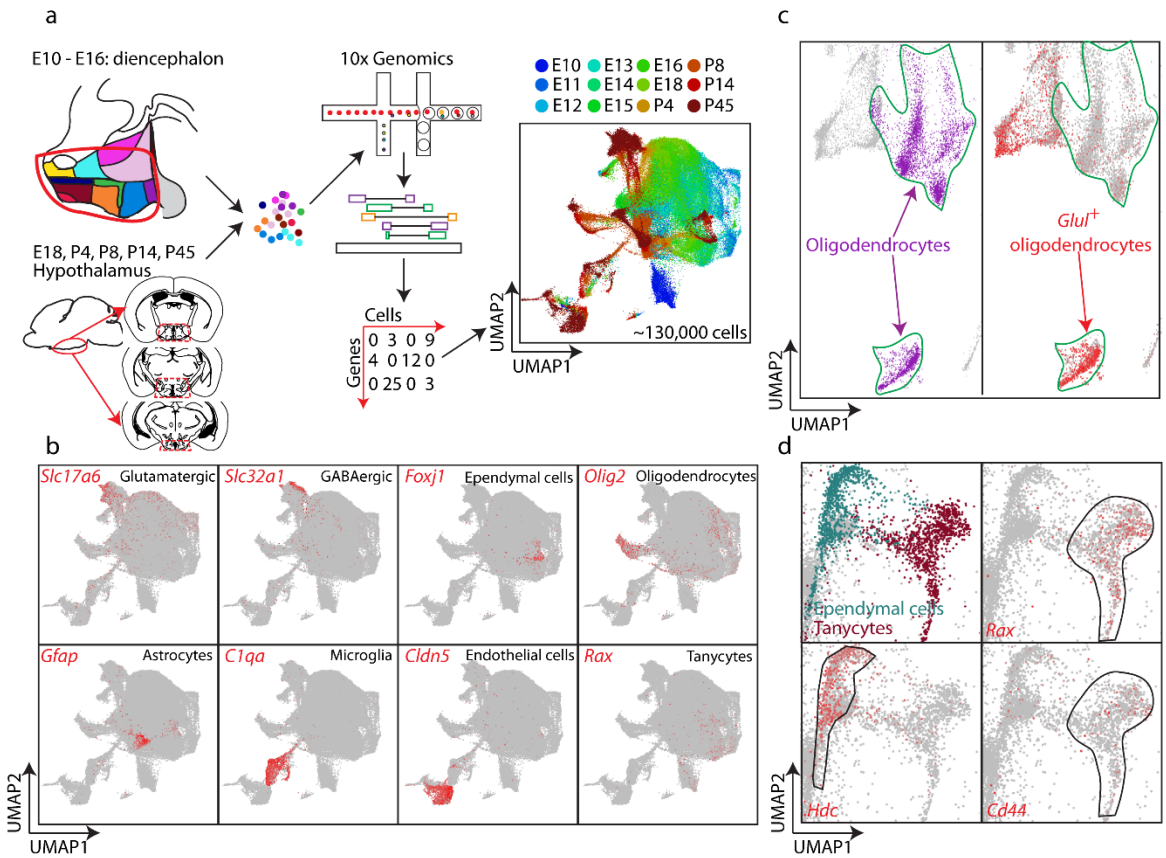

Fig1. Kim, et al.

Figure 1. Overview of generation of the hypothalamus scRNA-Seq dataset. (a) Schematic diagram showing overall experimental strategy. 12 time points of developing diencephalon including the prethalamus and hypothalamus (between E10 and E16), and hypothalamus (between E18 and P45) were profiled using the 10x Genomics Chromium system. Distribution of individual ages (blue = younger time point, red = older time point) is shown in the UMAP plot. (b) UMAP plot showing distribution of major cell types (red) of the hypothalamus at the terminal neuronal branch. (c) UMAP plot (higher power view) showing two separate populations of oligodendrocytes, *Glul*-positive and *Glul*-negative oligodendrocytes in the hypothalamus. (d) UMAP plot (higher power view) showing diverging developmental trajectories giving rise to leading to ependymal cells (green, *Hdc*) and tanycytes (brown, *Rax*, *Cd44*).

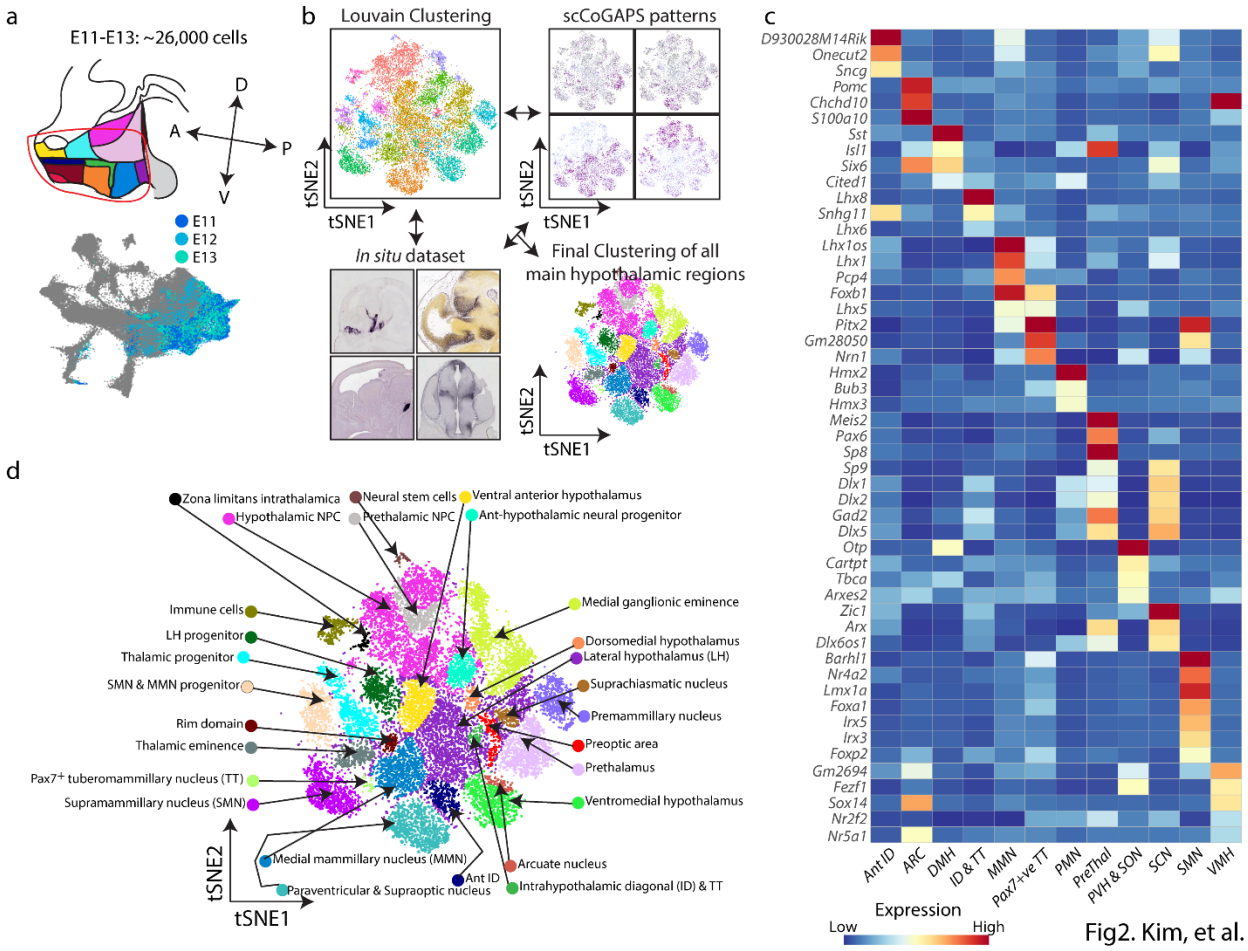

Figure 2. Specification of hypothalamic patterning during embryonic development. (a) Schematic diagram showing extraction of E11, E12 and E13 data to perform detailed analysis on hypothalamic patterning during development. (b) tSNE plot showing major subdivisions of the developing diencephalon (prethalamus and hypothalamus) and nearby regions to the developing diencephalon that are derived by cross-validation using the Louvain clustering algorithm, patterns from scCoGAPS and *in situ* analysis. (c) Heatmap showing a key subset of pattern-specific genes in major hypothalamic regions and prethalamus. (d) tSNE plot of E11-E13 developing diencephalon and adjacent regions.

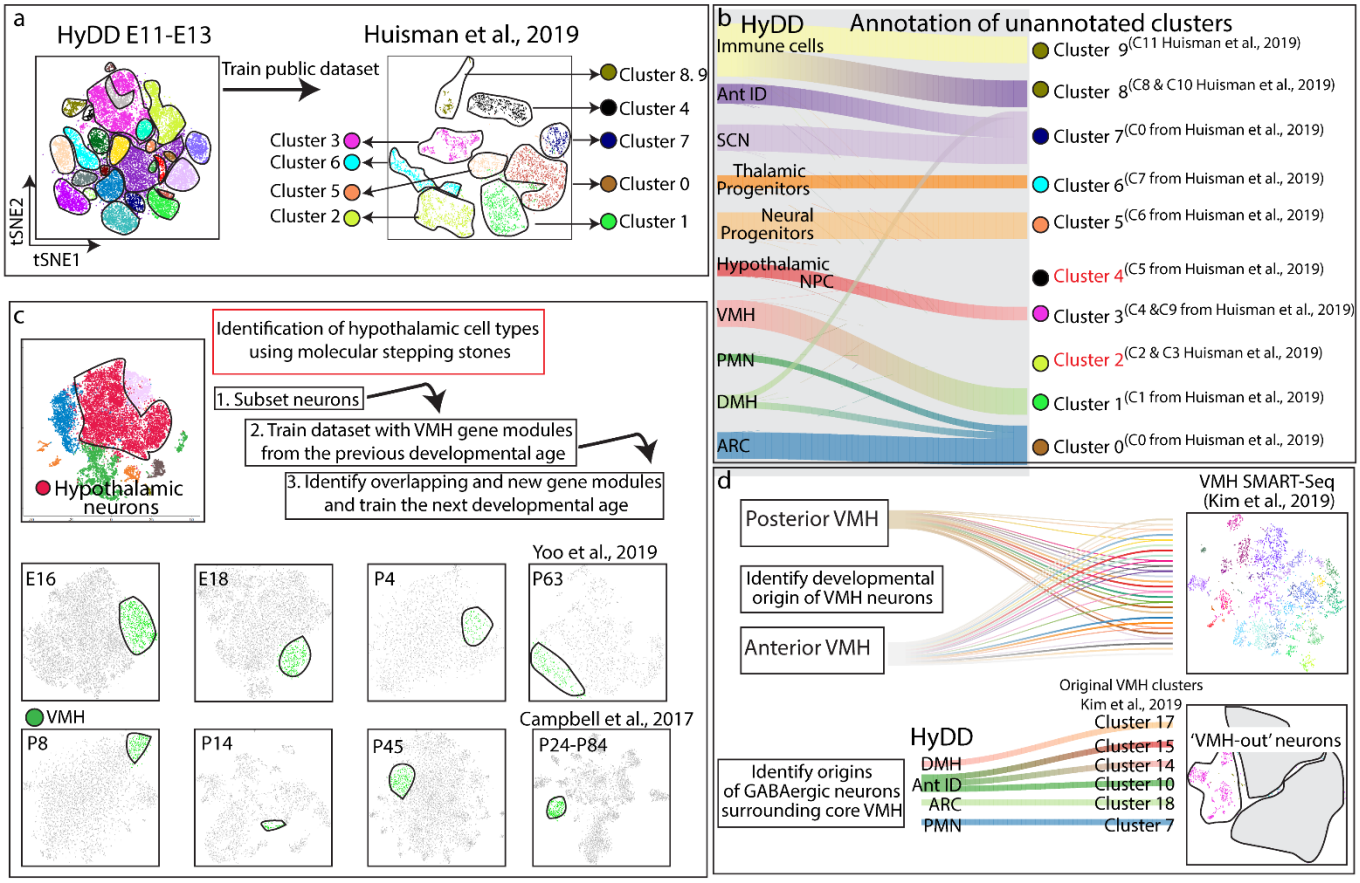

Fig3. Kim, et al.

Figure 3. Utilizing HyDD to infer the identity and origin of individual cell types in the developing and adult hypothalamus. (a) The HyDD dataset was used to train a previously published scRNA-Seq on E15.5 hypothalamus obtained through selective dissection of *Pomc-EGFP*-expressing cells[48]. (b) Alluvial plot showing HyDD clusters (left) matched to clusters from Huisman et al., 2019[48] (right). Note that 2 clusters (clusters 2 and 4) from Huisman et al., 2019[48] do not match the HyDD dataset. (c) Using the molecular stepping stone approach to identify VMH neurons (green) across the entire developmental ages by identification of shared sets of gene modules that can demarcate the VMH across the entire hypothalamus scRNA-Seq dataset. (d) HyDD dataset is used to identify the developmental origins of previously annotated subtypes of glutamatergic neurons of the core VMH[51] (top), and to identify the developmental origins of GABAergic neurons surrounding the core VMH (bottom).

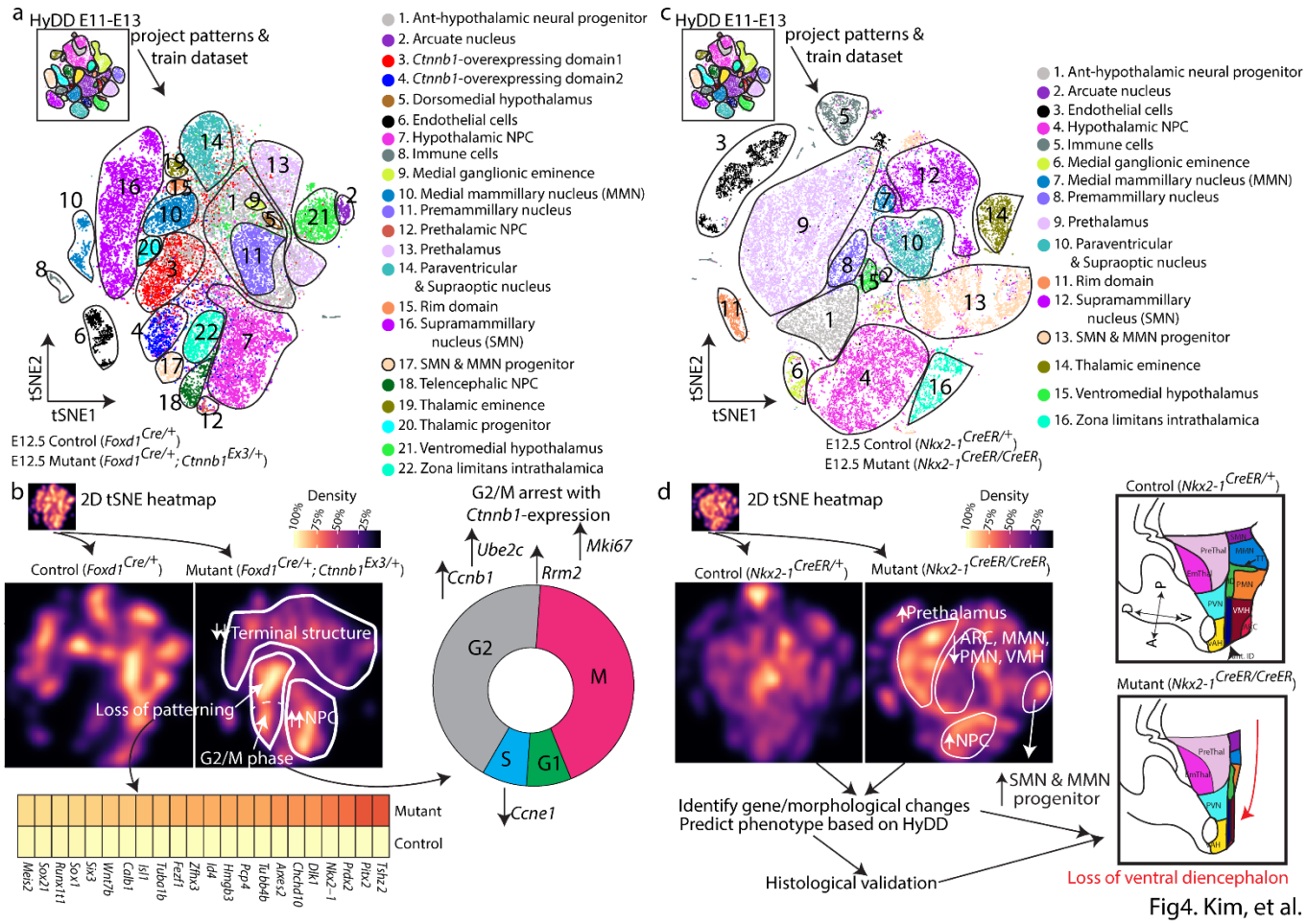

Figure 4. scRNA-Seq can be used to screen mutant phenotype in the hypothalamus. (a) tSNE showing distribution of diencephalic and nearby brain structures in control (*Foxd1*$^{Cre/+}$) and mice selectively overexpressing a constitutively active mutant form of *Ctnnb1* in neuroepithelial cells of the hypothalamus and prethalamus (*Foxd1*$^{Cre/+}$*;Ctnnb1*$^{Ex3/+}$) with clusters obtained by training the dataset with HyDD markers. Note distinct clusters (clusters 3 and 4) that are only observed in mutant samples)[20]. (b) tSNE heatmap showing distribution of individual cluster between control (*Foxd1*$^{Cre/+}$) and constitutively active *Ctnnb1* mutants (*Foxd1*$^{Cre/+}$*;Ctnnb1*$^{Ex3/+}$)(top left) and heatmap showing that mutant-specific clusters display all pattern-specific markers that are expressed across the developing diencephalon (bottom left), with a high level of G2/M phase markers in the mutant-specific cluster (right). (c) tSNE showing the distribution of hypothalamic, prethalamus and adjacent brain structures of control (*Nkx2-1*$^{CreER/+}$) and *Nkx2-1* mutant line (*Nkx2-1*$^{CreER/\ CreER}$) with clusters obtained by training the dataset with HyDD markers. (d) tSNE heatmap showing distribution of individual clusters between control (*Nkx2-1*$^{CreER/+}$) and *Nkx2-1* mutant line (*Nkx2-1*$^{CreER/\ CreER}$)(left). Note the absence of ventral diencephalic structures (except the supramammillary nucleus), and the relative expansion of the prethalamus in *Nkx2-1* mutants (right).