Midbrain dopamine (mDA) neurons play a central role in reward signaling and are widely implicated in psychiatric and neurodegenerative disorders. To understand how mDA neurons perform these functions, it is important to understand how mDA-specific genes are regulated. However, cellular heterogeneity in the mammalian brain presents a major challenge to obtaining this understanding. To this end, we developed a virus-based approach to label and capture mDA nuclei for transcriptome (RNA-Seq), and low-input chromatin accessibility (liDNase-Seq) profiling, followed by predictive modeling to identify putative transcriptional regulators of mDA neurons. Using this method, we identified Gmeb1, a transcription factor predicted to regulate expression of Th and Dat, genes critical for dopamine synthesis and reuptake, respectively. Gmeb1 knockdown in mDA neurons resulted in downregulation of Th and Dat, as well as in severe motor deficits. This study thus identifies Gmeb1 as a master regulator of mDA gene expression and function, and provides a general method for identifying cell type-specific transcriptional regulators.
Midbrain dopamine (mDA) neurons account for less than 1% of all neurons in the brain. Despite their limited number, mDA neurons are critical regulators of reward, cognition, and movement, and play a fundamental role in the pathophysiology of psychiatric and neurodegenerative disorders. In contrast to the well-defined role of dopamine signaling in health and disease, relatively little is known about how chromatin architecture regulates mDA gene expression. This knowledge gap is largely due to the cellular heterogeneity in the brain which impedes isolation of pure mDA nuclei from neighboring cell types, and therefore, generation of an mDA-specific map of accessible chromatin from which to identify transcriptional regulatory elements.

Chromatin accessibility is positively correlated to transcriptional activity. Actively transcribed genes often possess accessible (open) chromatin regions at their promoters and enhancers, thus permitting the recruitment of transcriptional factors and co-activators, whereas genes associated with inaccessible (closed) chromatin are generally transcriptionally repressed. As such, chromatin accessibility profiling has been used to gain insights into transcriptional regulation of various abundant cortical cell types during brain development.

In this study, we aim to understand how gene expression is regulated in mDA neurons, a limited, but important cell type. To this end, we developed a simple virus-based system to purify mDA nuclei from adult mouse brain for transcriptome and low-input chromatin accessibility profiling (Fig. 1a), and post a predictive model of transcriptional regulation. As a proof of principle, we focus on one of the candidate transcription factors, Gmeb1. We provide evidence demonstrating that Gmeb1 plays an important role in regulating the expression of Th and Dat, genes critical for dopamine neuron function, and show that Gmeb1 is necessary for maintenance of motor coordination.

Results
Cre-inducible HA nuclear tagging facilitates purification of mDA nuclei. To achieve mDA neuron-specific labeling, we designed a Cre-inducible AAV vector encoding the nuclear envelope protein KASH7,8 with an HA tag (Fig. 1b, top panel). To test whether the infection is specific to mDA neurons, we injected the AAV-DIO-KASH-HA (KASH-HA) virus into the midbrain of dopamine transporter (Dat)-Cre heterozygous (het) mice. Dat-Cre mice have been routinely used for targeted gene expression in mDA nuclei.

Gmeb1 serves as an mDA marker, revealed that 87.2% of mDA neurons expressed in HA+ mice were HA positive (HA+, over 40%) of these accessible chromatin sites were located in promoter regions, while HA+ nuclei were enriched for mDA-related functions such as response to cocaine and DA neuron differentiation, while HA− nuclei were enriched in non-neuronal functions such as gliogenesis, glial cell differentiation, and axon ensheathment (myelination), further confirming the purity of the sorted HA+ nuclei (Fig. 1e, Supplementary Fig. 1b).

Multi-omics analysis predicts Gmeb1 as an mDA transcriptional regulator. The mDA transcriptome presented in Fig. 1d, while depleted of glial gene expression, still included genes commonly found in other neuron types. To identify genes highly enriched in mDA neurons, we first derived a consensus mDA transcriptome from RNA-Seq of two biological replicates of purified mDA nuclei (Supplementary Fig. 2a) and then compared their expression level against three cortical neuron subtypes: vasoactive intestinal protein (VIP)-cortical, excitatory-cortical (Exc)-cortical, and parvalbumin (PV)-cortical neurons. This analysis revealed that out of the 394 HA+ genes, 107 are mDA-enriched (at least 4-fold higher in mDA neurons compared with VIP, Exc, and PV neurons, and q-value < 0.001) (Fig. 2a, b, Supplementary Fig. 2b, Supplementary Data 1). As expected, the list included mDA signature genes Th, Dat, Ddc, and Vmat. However, some of the HA+ enriched genes identified in Fig. 1d, such as Nurr1 and Satb1, were also highly expressed in cortical neuron subtypes and were thus excluded from the list of mDA-enriched genes (Supplementary Data 2). GO term and pathway analysis of the 107 genes further confirmed their enrichment in mDA neuron-specific functions such as dopamine metabolism, transport, and secretion (Supplementary Fig. 2c).

We performed RNA−seq on both the HA+ and HA− nuclei. Transcriptome analysis revealed that of the 13,727 genes expressed in HA+ and HA− nuclei (FPKM > 1), 394 genes exhibited at least a 4-fold enrichment in the HA+ cells, including known dopaminergic marker genes such as Th, Dat, Ddc, and Vmat. Conversely, 953 genes were significantly depleted (≥4-fold) in HA+ relative to HA− nuclei—including glial markers Gjb8 and Clcn1 (11). Gene ontology (GO) analysis (cutoff: q < 0.05) revealed that HA− nuclei were enriched for mDA-related functions such as response to cocaine and DA neuron differentiation, while HA+ nuclei were enriched in non-neuronal functions such as gliogenesis, glial cell differentiation, and axon ensheathment (myelination), further confirming the purity of the sorted HA+ nuclei (Fig. 1e, Supplementary Fig. 1b).

To this end, we performed liDNase-Seq 12, a technique that allows genome-wide identification of transcriptional regulatory elements using limited cell numbers. Using two biological replicates of 500 mDA nuclei (Supplementary Fig. 3a), we mapped the mDA DNasel-hypersensitive site (DHS) landscape. Among the 28,084 detected DHSs (p < 0.05), a large portion (over 40%) of these accessible chromatin sites were located in gene promoter regions (within ± 3 kb from transcription start site-TSS) (Supplementary Fig. 3b), consistent with previous reports in other cell types. By comparing mDA neuron DHS sites with the chromatin accessible sites of cortical neurons, we identified 2374 “mDA-enriched DHSs” open chromatin sites present in mDA neurons but absent from cortical neurons (Fig. 2c). Interestingly, the majority of the “mDA-enriched DHSs” were localized in non-promoter regions (3 kb away from TSS) (Supplementary Fig. 3c), highlighting the importance of distal DHSs in defining cell identity, in agreement with previous observations. Consistent with the positive correlation between chromatin accessibility and transcription, genes containing promoter DHSs in mDA neurons showed significantly higher expression levels compared with the genes lacking promoter DHSs (p-value = 2.2e−16, two-tailed Mann–Whitney–Wilcoxon Test), and this correlation was also maintained in mDA-enriched genes (p-value = 0.005397, two-tailed Mann–Whitney–Wilcoxon Test) (Supplementary Fig. 3d).
Out of the 107 mDA-enriched genes, 59 contained promoter DHSs. To identify candidate transcription factors (TFs) that could regulate these mDA-enriched genes, we performed genomic sequence motif enrichment analysis at promoter DHSs and identified 11 TF-binding motifs (p-value < 0.0001, FPKM > 1) (Fig. 2d, Supplementary Fig. 3e). This analysis not only identified known regulators of mDA gene expression such as Clock and Creb, but also Gmeb1 (glucocorticoid modulatory element binding protein-1). To ensure that Gmeb1 was not an artifact of transcriptome pruning, we performed a similar analysis using the 394 HA+ enriched gene promoters, and it also identified Gmeb1 (Supplementary Fig. 4b). Interestingly, Gmeb1 is a TF not previously known to play a role in mDA gene expression, whose binding motif is not present among cortical neuron-enriched
promoters (Supplementary Fig. 4a, c–e, Supplementary Data 3) despite its expression in cortical neurons5. Surprisingly, Nurr1 and Foxa2, two TFs extensively studied for their role in mDA neuron development and function18,19, were absent from mDA neuron DHSs (Fig. 2e, Supplementary Fig. 5, Supplementary Data 4).

To better understand the potential role of each of the 11 TF candidates in regulating the 59 mDA-enriched genes with accessible promoters, we assigned a “regulatory score” to each TF-gene pair where a higher score reflects a greater likelihood that the given TF would have the potential to regulate a given target gene through promoter DHS binding (Fig. 2f). To generate this predictive model, we designed four regulatory scoring schemes (Eqs. 1–4 in Methods) that associate gene expression with different TF-binding features to predict the regulatory effect for each of the 11 TFs on the expression level of the 59 genes. For each model we calculated the prediction error relative to the true expression level of the 59 genes (data used in Fig. 2b). We found that the most predictive scheme is the one that considered the TF-gene pair where a higher score reflects a greater likelihood that the given TF would have the potential to regulate a given target gene through promoter DHS binding (Fig. 2f). To generate this

![Heatmap showing the predicted TF regulatory score (TFRS) associating each TF to the mDA-enriched genes with a promoter DHS](image-url)
**Gmeb1 regulates transcription of mDA identity genes.** Gmeb1 has been shown to increase sensitivity to low glucocorticoid concentrations by acting as a transcription factor at the tyrosine transaminase promoter\(^{20}\), and has also shown to be a neuroprotective factor against oxidative stress\(^{21}\), but its role in mDA neuron function has not been previously implicated. Considering that the Gmeb1 binding motif is present in 35% of the accessible promoters of mDA-enriched-specific genes (Fig. 2d) and that two of the key mDA genes, Th and Dat, are predicted to be its targets, we hypothesized that Gmeb1 plays an important role in mDA neuron function. To test the transcriptional effects of Gmeb1 knockdown in mDA neurons, we designed three shRNAs targeting Gmeb1 and assessed their knockdown efficiency in N2A cells (Supplementary Fig. 7a, b). We then packaged the most efficient shRNA (sh1) into a vector (AAV-DIO-KASH-GFP-U6-shRNA), and delivered it into the midbrain of Dat-Cre mice (Fig. 3a). Two weeks after injection, midbrain tissue was dissected, and nuclei were isolated and immunostained for GFP. GFP\(^{+}\) (mDA) nuclei were FANS sorted and used for RNA-Seq to assess the transcriptional effects of Gmeb1 depletion. Transcriptome analysis of two biological replicates of control (shScramble) and Gmeb1 knockdown samples demonstrated high reproducibility (Supplementary Fig. 7c). Gmeb1 knockdown resulted in downregulation and upregulation (FC > 2) of 99 and 78 genes, respectively, in mDA neurons (Fig. 3b, Supplementary Data 6 and 7). The down-regulated genes included 9 mDA-enriched genes (Dat, Th, Cnp1, Agrp1a, Gyc2c, Aldh1a1, Ndnf, Anx11, Chnna). Notably, both Th and Dat, whose promoters contain Gmeb1 binding motifs, were significantly down-regulated following Gmeb1 knockdown (Fig. 3c). Immunostaining further confirmed depletion of Th and Dat at the protein level, in both midbrain and dorsal striatum, two principal projection regions for mDA neurons (Fig. 3d, e, Supplementary Fig. 8). Interestingly, a weak KASH signal was also detected in Th\(^{-}\) neurons in the shScramble group, suggesting that the vector may also be expressed in some non-mDA cells (Fig. 3d). This could be attributed to AAV design, where the knockdown vector (shScramble and shGmeb1) (Fig. 3a) contained a different backbone and promoter than the one used for nuclear tagging (Fig. 1b).

To ensure that the lack of Th or Dat protein signal was not due to cell death resulting from Gmeb1 knockdown, we compared the transcriptome of mDA neurons 2 weeks after infection with viruses expressing shGmeb1 or shScramble. We found that none of the 7 queried apoptosis-related genes showed significant alteration in Gmeb1 knockdown mDA neurons compared with the shScramble controls (Supplementary Fig. 9a). This result is consistent with the lack of significant increase in the number of cleaved caspase-3 positive mDA neurons (Supplementary Fig. 9b, c). However, these results cannot rule out the possibility that inflammatory processes may induce cell death by necrosis, or that cells may have already degenerated through apoptosis at this timepoint, leaving only the surviving cells for transcriptome analysis.

To address these possibilities, we first assessed the expression levels of genes involved in necrosis and found no significant alteration in response to Gmeb1 knockdown (Supplementary Fig. 9d). Next, we co-injected viruses expressing either shGmeb1 and cre-inducible mCherry (DIO-mCherry), or shScramble and DIO-mCherry into the substantia nigra pars compacta (SNC) of Dat-Cre mice. This approach ensured that mDA neurons could be identified and counted despite Th signal loss following Gmeb1 knockdown. We found that Gmeb1 knockdown in the SNC, while resulting in Th loss (Supplementary Fig. 10a), does not change the number of mDA neurons (mCherry\(^{+}\)), thus suggesting no mDA neuron loss (Supplementary Fig. 10c). This inference was further supported by TUNEL assay, in which fragmented DNA, a hallmark of cell death, was rarely detected in either shGmeb1 or shScramble group (Supplementary Fig. 10a). In contrast, treatment with DNase-1, which induces DNA fragmentation, resulted in robust TUNEL signal (Supplementary Fig. 10a), suggesting that loss of Gmeb1, while abrogating Th and Dat expression, does not induce cell death.

While Gmeb1 knockdown does not result in mDA neuron death, the transcriptional consequence of this manipulation could alter the basic electrophysiological properties of mDA neurons due to its effect on dopaminergic transmission. To explore this possibility, we tested the effect of Gmeb1 knockdown on the excitability of SNC mDA neurons by current clamp, using the viral injection strategy described in Supplementary Fig. 10b. We found that loss of Gmeb1 did not significantly change the nature or frequency (Hz) of evoked action potentials (Fig. 4a, Supplementary Fig. 11) when compared with shScramble controls, nor the amplitude (mV) of the recorded potentials (Fig. 4b, Supplementary Fig. 11a). Furthermore, hyperpolarization currents resulted in down regulated sag potentials in both groups (Supplementary Fig. 11b). Collectively, these results suggest that loss of Gmeb1 does not affect SNC mDA neuron excitability, firing rate or magnitude of evoked action potentials.

**Gmeb1 is required for maintaining homeostatic motor coordination.** Tyrosine hydroxylase (Th) is the rate-limiting enzyme in catecholamine synthesis\(^{22}\) and therefore depletion of Th results in loss of dopamine\(^{23}\). Dysregulation of dopamine signaling in humans has been associated with mood disorders, drug addiction and is the root cause of motor impairments associated with Parkinson’s disease (PD)—a neurodegenerative disorder characterized by progressive loss of nigral mDA neurons, resulting in reduced motor coordination, balance and increased muscle fatigue.\(^{2,3}\) Given that Gmeb1 plays a critical role in regulating Th expression, we next asked whether Gmeb1 would be necessary to maintain normal motor functions. To this end, we tested mice with bilateral SNC Gmeb1 knockdown in a battery of motor assessments, including the pole test, rotarod test, swim test and hanging wire test. In the pole test, a mouse is required to walk down a 50-cm grooved vertical pole from top to base, and animals with balance impairments will take longer to reach the base of the pole. We found that SNC Gmeb1-knockdown mice (shGmeb1) took more time to reach the base than control (shScramble) mice (Fig. 5a, Supplementary Movie 1; n = 11/group, ***p < 0.0001 with two-tailed t-test). Consistent with pharmacological models of dopamine depletion\(^{24,25}\), our results suggest that Gmeb1 knockdown in SNC results in balance and coordination impairments.

The rotarod test, which requires mice to continuously walk on an accelerating rod over the span of 5 min is also used to measure balance and coordination\(^{26}\) and the “latency”—the amount of time the mouse stays on the rod without falling or clinging on, is recorded. We found that Gmeb1 knockdown mice exhibited lower latency than control mice (Fig. 5b, Supplementary Movie 2, n = 11/group, ***p < 0.0001 with two-tailed t-test). This result indicates that Gmeb1 knockdown impairs rotarod performance at lower speeds, and thus suggests a deleterious effect on balance and coordination, consistent with results of the pole test (Fig. 5a). However, since climbing and trotting are behaviors in which nearly all experimental mice have daily experience, we tested the animals’ ability to adapt to a new environment by performing an innate behavior in which they had no prior experience. To this end, we placed mice in a water-filled chamber and assessed the animals’ swimming ability over the span of 5 min. Using a scale of 0–5 where the higher score indicates greater swimming
proficiency, Gmeb1 knockdown mice showed almost complete inability to swim (Fig. 5c, n = 11/group, ****p < 0.0001 with two-tailed t-test). In contrast, when the shScramble mice were placed in the water-filled chamber, they adapted their motor repertoire and swam for at least 5 min while Gmeb1 knockdown mice generally floated with their hind limbs spread horizontally and struggled to coordinate the hind limb movement necessary to swim (Supplementary Movie 3). These results suggest that Gmeb1

Fig. 3 Midbrain Gmeb1 knockdown results in loss of Th and Dat. a Diagram of the DIO-KASH-GFP-shRNA expressing vector (top) and location of midbrain virus injection in Dat-Cre mice (bottom). b Scatter plot comparing RNA-Seq data of mDA neurons of control (shScramble, n = 2) and Gmeb1 knockdown (shGmeb1, n = 2). Gmeb1 knockdown results in down-regulation of 99 genes (blue), including Th and Dat, and up-regulation of 78 genes (red) (fold change > 2; p-value < 0.05). c Histograms (RPKM) and genome browser view showing down-regulation of Th and Dat following Gmeb1 knockdown. Dashed inserts: Location of Gmeb1 motifs in promoter DHS for Th and Dat (red triangles). d Representative immunostaining pictures showing KASH-HA tagged mDA nuclei (red) and reduction of Th (green) in VTA of shGmeb1-treated mice. Box inserts reflect brain region where micrograph was obtained. Scale bar: 500 μm. Note: Coronal brain section was reproduced from ref. 48. (Copyright 2013, Elsevier, Academic Press). e Representative immunostaining pictures showing reduction of Dat (green) in caudate putamen (CPu) of shGmeb1-treated mice. LV; Lateral ventricle. Box inserts reflect brain region where micrograph was obtained. Scale bar: 500 μm. Note: Coronal brain section was reproduced from ref. 48. (Copyright 2013, Elsevier, Academic Press)
knockdown results in motor deficits, especially when animals are challenged with a situation that requires a new form of motor response.

Lastly, to determine if muscle endurance contributed to the above Gmeb1 loss-of-function phenotypes, we subjected the mouse cohorts to the hanging wire test. In this test, mice grasp a 2-mm diameter steel wire 40 cm above a padded surface so that the animal hangs by gripping the wire with its forepaws. The time (latency), required for the mouse to fall is recorded, where lower latency to fall indicates lower muscle endurance. We found no significant difference between the groups (Fig. 5d, Supplementary Movie 4, n = 11/group, p = 0.3462 with two-tailed t-test), suggesting that muscle endurance is not affected, and that the effects seen in the pole test, rotarod and swim tests are most likely due to deficits in balance and/or coordination. While muscle endurance is compromised in PD, this phenomenon can also be attributed in part to muscle atrophy. Given that mice were tested approximately two weeks after virus injection, it is not surprising that this aspect of motor control was not yet affected by Gmeb1 knockdown. Further, knocking down Gmeb1 did not result in hypolocomotion, as measured by distance traveled (Supplementary Fig. 12). Taken together, this is the first study to generate a genome-wide chromatin accessibility map of mDA neurons. By comparing transcriptome and accessible chromatin maps of mDA neurons to those of cortical neurons, we identified candidate TFs predicted to regulate mDA-enriched gene expression (Fig. 2e). While our TF motif analyses identified known mDA transcriptional regulators, such as Nurr1 and Foxa2, it also uncovered novel transcriptional regulators, such as Gmeb1, whose function in mDA neurons was previously unknown.

As a proof-of-principle study, we demonstrated that Gmeb1 plays an important role in regulating mDA-enriched genes, such as Th and Dat, which are essential for dopamine signaling (Fig. 3). Indeed, Gmeb1 knockdown reduces Th and Dat expression, yet it does not affect mDA neuron survival (Supplementary Figs. 9, 10) or basic electrophysiological functions (Fig. 4). This suggests that while Gmeb1 knockdown may not affect the ability of mDA neurons to communicate, due to the essential role of Th in dopamine synthesis, their reduced Th levels may compromise homeostatic dopamine signaling. Consistent with this notion, knockdown of Gmeb1 in the SNC of adult mice resulted in motor deficits, similar to pharmacological models of PD (Fig. 5). Collectively, our study reveals Gmeb1 as a novel transcriptional regulator essential for mDA neuron function.

**Discussion**

Understanding transcriptional regulation of genetically-defined neuron populations in vivo has proven difficult due to the cellular heterogeneity of the mammalian brain. Despite efforts in profiling the mDA transcriptome using tools such as TRAP-Seq and single-cell RNA-Seq, these approaches do not allow for chromatin analysis, and therefore do not provide mechanistic insights into how mDA genes are regulated—especially those directly involved in dopamine signaling. To overcome these technical hurdles, we developed an in vivo virus-based approach to tag and purify mDA nuclei for transcriptome and chromatin accessibility analysis (Fig. 1a).

mDA neurons play a central role in reward signaling and are widely implicated in psychiatric and neurodegenerative disorders. Therefore, a better understanding of transcriptional regulation in this neuronal population is critical. To our knowledge, this is the first study to generate a genome-wide chromatin accessibility map of mDA neurons. By comparing transcriptome and accessible chromatin maps of mDA neurons to those of cortical neurons, we identified candidate TFs predicted to regulate mDA-enriched gene expression (Fig. 2e). While our TF motif analyses identified known mDA transcriptional regulators, such as Nurr1 and Foxa2, it also uncovered novel transcriptional regulators, such as Gmeb1, whose function in mDA neurons was previously unknown.

As a proof-of-principle study, we demonstrated that Gmeb1 plays an important role in regulating mDA-enriched genes, such as Th and Dat, which are essential for dopamine signaling (Fig. 3). Indeed, Gmeb1 knockdown reduces Th and Dat expression, yet it does not affect mDA neuron survival (Supplementary Figs. 9, 10) or basic electrophysiological functions (Fig. 4). This suggests that while Gmeb1 knockdown may not affect the ability of mDA neurons to communicate, due to the essential role of Th in dopamine synthesis, their reduced Th levels may compromise homeostatic dopamine signaling. Consistent with this notion, knockdown of Gmeb1 in the SNC of adult mice resulted in motor deficits, similar to pharmacological models of PD (Fig. 5). Collectively, our study reveals Gmeb1 as a novel transcriptional regulator essential for mDA neuron function.

**Fig. 4** Gmeb1 knockdown does not impair electrophysiological function in SNC mDA neurons. **a** Sample current-clamp recordings of action potentials from SNC mDA neurons following injection of depolarizing currents (pA). **b** Gmeb1 knockdown does not affect the mean plateau voltage or AP frequency of SNC mDA neurons. Shown are the mean (±s.e.m.) plateau voltage (mV) curves (left) and the mean (±s.e.m.) action potential frequency (Hz) (right) of Gmeb1 knockdown (n = 26 cells) or control (n = 19 cells) SNC mDA neurons following incremental current (pA) injection (right).

**Fig. 5** Knockdown of Gmeb1 does not impair motor function. **a** Knockdown of Gmeb1 does not impair motor function as measured by distance traveled (Supplementary Fig. 11). **b** Knockdown of Gmeb1 does not affect latency to fall in the hanging wire test (Supplementary Fig. 12).
Methods

**Animals.** Female Dat-IRES-Cre heterozygous knock-in mice (Jackson Laboratories, 06660) were bred with male C57BL/6 wildtype mice (Jackson Laboratories, 006664) to produce Dat-IRES-Cre heterozygous and wildtype offspring. Only male mice were used for profiling experiments whereas both male and female mice were used for histological analysis and establishment of the infection system. Mice were 8–12 weeks of age at the beginning of each experiment. All animal husbandry and behavioral procedures were conducted in strict accordance with the NIH Guide for the Care and Use of Laboratory Animals and were approved by the Institutional Animal Care and Use Committee of Harvard Medical School.

**Genotyping.** Around 21 days of age, mouse pups were weaned and their ears were clipped for genetic analysis. Genomic DNA was extracted using the hot sodium hydroxide and tris (HotSHOT) method. Primers for the Dat wildtype and Dat-Cre mutant genes were: olMR6625—common (5′-TGGCTGTTGGTGTAAAGT TT-3′); olMR6626—wildtype reverse (5′-GGCCAGACATTGTTGACT-3′), DAT-Cre—mutant reverse (5′-CCTAAGGATGGCAATGTGGT-3′). Samples were processed for genetic amplification with PCR and analyzed on a 1.5% agarose gel with GelRed Nucleic Acid Stain. The band for the DAT-Cre wildtype gene was at 264 bp, and the Dat-Cre mutant gene was at 152 bp.

**Vector construction.** To generate the AAV-DIO-KASH-HA vector, KASH sequence (kind gift from Dr. Feng Zhang, Broad Institute) was cloned into the DIO cassette of cDIO-L10-VHII-HA (kind gift from Dr. Jeffrey Friedman, Rockefeller University), replacing L10-VHII sequence. To generate the AAV-DIO-KASH-GFP-U6-shRNA vector, the EGFP sequence of the AAV-hsyn-cDIO-EGFP vector (Addgene 50457) was replaced by KASH-GFP sequence and then the U6-shRNA vector of the AAV-hsyn-cDIO-shRNA vector (Addgene 50460) was inserted after the KASH-GFP sequence. The KASH sequence (kind gift from Dr. Feng Zhang, Broad Institute) was cloned into the DIO expression vector. The KASH sequence was inserted after the polyA sequence. Using aseptic technique, a 5 mm long transverse incision was made on the skin overlying the skull, exposing Bregma. Two stereotaxic surgery and viral delivery. Animals were anesthetized with a 1–3% isoflurane/oxygen mixture and mounted in a stereotactic frame (Kopf Instruments, Tujunga, CA) at a “flat-skull” position. Using aseptic technique, a 5 mm longitudinal incision was made on the skin overlying the skull, exposing Bregma. Two
small circular openings were drilled on the skull to expose the dura surface overlying the midbrain. Two bilateral injections (0.5 µl each at a flow rate of 0.2 µl per min) were made using the following coordinates: For midbrain, anterior-posterior (AP) = −2.95 mm; mediolateral (ML) = ± 0.5 mm from midline; dorsal-ventral (DV): −4.2 mm from dura. For SNc, AP = −2.95 mm; ML ± 1.4 mm from midline; DV −4.3 mm from dura. To ensure proper viral dispersion throughout midbrain parenchyma, the 30-gauge needle was left in place for 5 min before retracting.

**Tissue dissection, brain perfusion, and fixation.** For RNA-Seq and Immunohisto-Seq, mice were euthanized by inhalation of CO2. Brains were rapidly removed and the remaining tissue was perfused through the ascending aorta with 0.9% saline, followed by 4% paraformaldehyde in 0.125 M phosphate buffer (pH 7.4). Tissues were stored in 30% sucrose in 0.1 M phosphate buffer (pH 7.4) at 4 °C. After removing the supratentorium, brains were cut into 200 µm sections and placed in wash buffer (1% BSA, 150 mM NaCl, 5 mM MgCl2, 0.1% Triton X-100) and incubated on ice for 5 min. The tissues were washed with wash buffer for 1 h and centrifuged (5 min at 1000 × g) at room temperature. After removing the supernatant, the sections were incubated in wash buffer containing (all secondary antibodies were used at 1:500 concentration): mouse anti-Caspase-3 (1:500, Ab-13847, Abcam), rat anti-Dat (1:500, SC-25269, Santa Cruz), rabbit anti-HA (1:800, 3724, Cell Signal- ing), Chicken anti-GFP (1:2500, Ab-13970, Abcam) and rat anti-Th (1:500, SC-25269, Santa Cruz). The DNA was eluted in TE. The libraries were sequenced on a Hiseq2500 with single-end 100 bp reads (Illumina). The DNA was eluted in TE. The libraries were sequenced on a Hiseq2500 with single-end 100 bp reads (Illumina). The DNA was eluted in TE. The libraries were sequenced on a Hiseq2500 with single-end 100 bp reads (Illumina). The DNA was eluted in TE. The libraries were sequenced on a Hiseq2500 with single-end 100 bp reads (Illumina). The DNA was eluted in TE. The libraries were sequenced on a Hiseq2500 with single-end 100 bp reads (Illumina). The DNA was eluted in TE. The libraries were sequenced on a Hiseq2500 with single-end 100 bp reads (Illumina). The DNA was eluted in TE. The libraries were sequenced on a Hiseq2500 with single-end 100 bp reads (Illumina). The DNA was eluted in TE. The libraries were sequenced on a Hiseq2500 with single-end 100 bp reads (Illumina). The DNA was eluted in TE. The libraries were sequenced on a Hiseq2500 with single-end 100 bp reads (Illumina).
expression change were considered. For the Gmeb1 knockdown RNA-Seq data, a lower cutoff criterion (RPKM > 1, q-value < 0.05 and FC > 2) was used due to the variability between the knockdown samples.

**Functional enrichment analysis.** The enrichGO and enrichKEGG function from the R/Bioconductor clusterProfiler package were, respectively, used to perform GO and pathway enrichment analysis. Only the GO terms and pathways with an adjusted p-value < 0.05 were considered. The associated GO and pathway enrichment plots were generated using the ggplot2 package.

**Plot generation.** Heatmaps were generated using the R/Bioconductor package ComplexHeatmap. Boxplots were generated using the standard R boxplot function. All the other plots were generated using the ggplot2 package.

**DNase-seq data analysis.** Raw 100-bp single-end DNase-seq reads were first trimmed with Trimmomatic (v.0.36) with parameters similar to those used in RNA-Seq, except that trimmed reads were allowed to have a length of at least 25-bp. Next, we used Bowtie2 to map the trimmed reads to the mm9 genome without allowing any mismatches. Unmapped reads and reads with mapping quality less than 10 were filtered out using samtools (v.1.3.1). DNase-seq peaks were then called using macs2 with the parameters—broad—nomodel—nolambda. Peaks with a q-value < 0.05 were considered for further analysis.

**ATAC-seq analysis.** ATAC-seq reads for each replicate of Exc, PV and VIP neurons were mapped to the mm9 genome. Raw reads were first trimmed using Trimmomatic with the same parameters used for idDNase-seq analysis. The trimmed paired-end reads were then mapped using Bowtie2 with parameter-X 2000. Unmapped reads, reads with mapping quality less than 10 and reads mapping to the mitochondrial genome were discarded. Macs2 was then used with parameters—broad—nomodel—nolambda to call peaks. Peaks with a q-value < 0.05 were considered.

**Peaks genomic annotation.** NCBI/UCSC peak16.76 was used as a source of gene annotation. Mainly, the annotatePeak function from the R/Bioconductor ChIPpeakAnno package was used for genomic annotation. Promoters were defined by 13kb from TSS of the transcripts, and all regions that did not fall within exons, introns, or UTRs were classified as distal intergenic regions.

**Promoter DHS motif analysis.** Motif analysis was performed using the R/Bioconductor PWMEnrich package. To find motifs enriched at the promoters of mDA-enriched genes, we first constructed a lognormal background motif distribution of the 200-bp chunks generated from the TSS ± 3 kb of all the annotated NCBI/UCSC mouse (mm9) genes. As a foreground, we used the mm9 DHS peaks located within 3 kb from the TSS of the TSS of the 107 mDA-enriched genes. A total of 59 genes out of 107 mDA-enriched genes contained a promoter DHS. The motifEnrichment function was used to detect the most enriched motifs.

**Distal DHS motif analysis.** To locate distal regulators of mDA neurons, we selected DHSs located 3 kb to 1 Mb away from TSS for each of the 107 mDA-enriched genes. Next, we constructed a lognormal background motif distribution of the 200-bp chunks generated from the non-specific distal mDA DHS that overlapped with ATAC-seq peaks in VIP, PV or Exc neurons. Then, the motifEnrichment function was used to find the enriched motifs. Because the most enriched distal motifs in our analysis were known mDA neurons regulators (Nurr1 and FoxA2), we adjusted the p-value to < 1e-2 and mRNA expression level to FPKM > 1 to reveal potential new distal regulators.

**Associating promoter-enriched TF with their target genes.** Motif analysis only produced a list of over-represented TF binding-site (TFBS) in promoter DHS, but did not directly indicate which TF would regulate which gene(s). To predict the potential of a given TF to regulate a gene, we calculated a TF regulatory score (TFRS) for each TF-gene pair. Initially, we defined four regression models (elastic net regression) by finding a correlation between gene expression and different TF binding features, then, we selected the model with the best predictive outcome. Three features were considered: (i) the binding affinity of the TF to the gene’s promoter DHS (represented by the TF motif p-value), (ii) the strength of the gene’s promoter DHS peak signal (RPKM) and (iii) the distance of the promoter DHS peak to the gene’s TSS. For each model, the 59-gene by 11-TF matrix was constructed in which the TFRS score of TF, to a gene g was calculated with one of the following model scores:

\[
TFRS^{\text{p-value}}_g = \log_{10} \left( \text{p-value}^{-1}_g \right) \\
TFRS^{\text{p-value}}_g = \log_{10} \left( \text{p-value}^{-1}_g \right) \times \text{RPKM}_{\text{DHS(g)}} \\
\]
session ended and the mouse returned to its home cage. The testing session was repeated 30 min later. Scores for each session were averaged to represent the composite score shown in Fig. 5k.

**Swim test.** C57BL/6 mice with either shGmeb1 or shScramble virus injection were subjected to the swim test. Mice were placed inside a 2 L beaker (13 cm diameter) filled with 1400 ml of water at 25°C and were allowed to swim for 5 min, during which time the animals’ swimming dexterity was scored. The swimming score (range 0–5) was assigned as follows: 5: continuous swimming (>80% of session); 4: continuous swimming (>60% of session) with occasional bouts of floating; 3: equal time (±10%) spent floating/swimming; 2: floating with occasional bouts of swimming (<20% of session); 1: Same as score of 2, but swimming bouts display limited hind limb paddling motion; 0: mice removed from the swim chamber before 5 min due to drowning. To prevent unnecessary distress, mice were removed from the chamber at the first sign of drowning. At the conclusion of the test, mice were removed, dried and returned to their home cage. Swim scores were averaged to represent the composite score shown in Fig. 5c.

**Hanging wire test.** C57BL/6 mice with either shGmeb1 or shScramble virus injection were subjected to the hanging wire test. Mice were placed on top of a steel wire 2 mm in diameter, suspended 40 cm above soft padding material. Placement of the mouse was such that both forepaws completely gripped the steel wire while the body hanged below. Once the mouse was placed on the steel wire, the time was recorded with a stopwatch until the mouse fell from the wire and landed on the soft padding surface. The amount of time the mouse hanged on the wire without falling was termed the “latency” score. The test was repeated twice with a 30-min interval between trials, during which the mouse was returned to its home cage. Latency scores from both trials were averaged to represent the composite score shown in Fig. 5d.

**Open-field locomotion.** C57BL/6 mice with either shGmeb1 or shScramble virus injection were subjected to open field locomotion test. Prior to testing, mice were habituated in the room for at least 30 min and then placed in the center of the open-field arena (Med Associates, ENV-510). The mouse was allowed to move freely in the arena for 30 min, which would be recorded as beam breaks and recorded as “distance traveled”. At the conclusion of the test, the mouse was removed and returned to its home cage. Distance traveled (cm) scores for each group were averaged to represent the composite scores shown in Supplementary Fig. 10.

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

**Data availability**

The accession number for the RNA-seq and idDNase-seq data presented in this study is available from the Gene Expression Omnibus (GEO) database under accession GSE106956.

**Code availability**

Additional custom codes used for bioinformatics analysis are available upon request.

Received: 24 April 2018 Accepted: 26 April 2019

**References**

1. Nelson, E. L., Liang, C. L., Sinton, C. M. & German, D. C. Midbrain dopaminergic neurons in the mouse: computer-assisted mapping. *J. Comp. Neurol.* 369, 361–371 (1996).
2. Hyman, S. E., Malenka, R. C. & Nestler, E. J. Neural mechanisms of addiction: the role of reward-related learning and memory. *Ann. Rev. Neurosci.* 28, 565–598 (2005).
3. Moore, D. J., West, A. B., Dawson, V. L. & Dawson, T. M. Molecular pathophysiology of Parkinson’s disease. *Ann. Rev. Neurosci.* 28, 57–87 (2005).
4. Gross, D. S. & Garrard, W. T. Nucleace hypersensitive sites in chromatin. *Annu Rev. Biochem.* 57, 159–197 (1988).
5. Mo, A. et al. Epigenetic signatures of neuronal diversity in the mammalian brain. *Neuron* 86, 1369–1384 (2015).
6. Preisel, S. et al. Single-nucleus analysis of accessible chromatin in developing mouse forebrain reveals cell-type-specific transcriptional regulation. *Nature Neurosci.* 21, 432–439 (2018).
7. Osozawa, C. et al. Dynamics and molecular interactions of linker of nucleoskeleton and cytoskeleton (LINC) complex proteins. *J. Cell Sci.* 122, 4099–4108 (2009).
8. Schwiehl, L. et al. In vivo interrogation of gene function in the mammalian brain using CRISPR-Cas9. *Nat. Biotechnol.* 33, 102–106 (2015).
9. Buxton, B. M. et al. Characterization of a mouse strain expressing Cre recombinase from the 3′ untranslated region of the dopamine transporter locus. *Genesis* 44, 383–390 (2006).
10. Poulin, J. F. et al. Defining midbrain dopaminergic neuron diversity by single-cell gene expression profiling. *Cell Rep.* 9, 930–943 (2014).
11. Caboy, J. D. et al. A transcriptome database for neurons, synapses, and oligodendrocytes: a new resource for understanding brain development and function. *J. Neurosci.* 28, 264–278 (2008).
12. Lu, F. et al. Establishing chromatin regulatory landscape during mouse preimplantation development. *Cell* 165, 1375–1388 (2016).
13. Song, L. et al. Open chromatin defined by DNase and FAIRE identifies regulatory elements that shape cell-type identity. *Genome Res.* 21, 1757–1767 (2011).
14. Thurman, R. E. et al. The accessible chromatin landscape of the human genome. *Nature* 489, 75–82 (2012).
15. Vierstra, J. et al. Mouse regulatory DNA landscapes reveal global principles of cis-regulatory evolution. *Science* 346, 1007–1012 (2014).
16. McClung, C. A. et al. Regulation of dopaminergic transmission and cocaine reward by the Clock gene. *Proc. Natl Acad. Sci. USA* 102, 9377–9381 (2005).
17. Andersson, M., Konradi, C. & Cenci, M. A. cAMP response element-binding protein is required for dopamine-dependent gene expression in the intact but not the dopamine-derivatized striatum. *J. Neurosci.* 21, 9930–9943 (2001).
18. Kedokhodei, B. et al. Nurr1 is required for maintenance of maturing and adult midbrain dopamine neurons. *J. Neurosci.* 29, 15923–15932 (2009).
19. Pristera, A. et al. Transcription factors FOXA1 and FOXA2 maintain reward by the Clock gene. *J. Neurosci.* 35, 187–197 (2015).
20. McElhinny, A. S., Greene, J. G. & Miller, G. W. Behavioral phenotyping of mouse models of Parkinson’s disease. *Behav. Brain Res.* 211, 1–10 (2010).
21. Miu, C., Torres, E. M. & Dunnett, S. B. Comparison of incremental and accelerating protocols of the rotarod test for assessment of motor deficits in the 6-OHDA model. *J. Neurosci. Methods* 158, 219–223 (2006).
22. Petroni, M. L. et al. Body composition in advanced-stage Parkinson’s disease. *Acta Diabetol.* 40(Suppl 1), S187–S190 (2003).
23. Ekstrand, M. L. et al. Molecular profiling of neurons based on connectivity. *J. Biol. Chem.* 280, 1230–1242 (2014).
24. Brichta, L. et al. Identification of neurodegenerative factors using translatome-regulatory network analysis. *Nat. Neurosci.* 18, 1325–1333 (2015).
25. La Manno, G. et al. Molecular diversity of midbrain development in mouse, human, and stem cells. *Cell* 167, 566–580 e519 (2016).
26. Draett, G. E. et al. Preparation of PCR-quality mouse genomic DNA with hot sodium hydroxide and tris (HotSHOT). *Biotechniques* 29, 54 (2000).
27. Wallace, M. L. et al. Genetically distinct parallel pathways in the entopineural nucleus for limbic and sensorimotor output of the basal ganglia. *Neuron* 94, 138–152 e135 (2017).
28. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120 (2014).
29. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29, 15–21 (2013).
30. Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *Bioinformatics* 20, 520–521 (2014).
31. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with Deseq2. *Genome Biol.* 15, 550 (2014).
32. Yu, G., Wang, L. G., Han, Y. & He, Q. Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* 16, 284–287 (2012).
33. Wickham, H. *ggplot2: elegant graphics for data analysis.* (Springer, New York 2009).
34. Zarei, Z., Ehs, R. & Schlesner, M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* 32, 2847–2849 (2016).
Acknowledgements

We thank Drs. Jeffrey Friedman (Rockefeller U) and Feng Zhang (Broad Institute) for providing plasmids; Drs. Wenqiang Chen and Aritra Bhattacharjee for critical reading of the manuscript; Barbara Caldarone and Paul Lorello (Harvard Neurobehavior Core) for providing logistical assistance with animal studies, and the reviewers of this manuscript for their time and constructive feedback. This project is supported by NIH R01DA042283 (to Y.Z.), NIH K01DA045294 (to L.M.T.) and a Klarman Family Foundation Eating Disorders Research grant (to Y.Z.). Y.Z. is an investigator of the Howard Hughes Medical Institute.

Author contributions

Y.Z. conceived the project; L.M.T., B.L.S., and Y.Z. designed the experiments and wrote the manuscript; L.M.T., M.N.D., R.C., F.L., W.W., and B.L.S. performed experiments and analyzed the data.

Additional information

Supplementary Information accompanies this paper at https://doi.org/10.1038/s41467-019-10267-0.

Competing interests: The authors declare no competing interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

Journal Peer Review Information: Nature Communications thanks Angel Barco, and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s) 2019