Transcriptional Profiling Reveals the Transcription Factor Networks Regulating the Survival of Striatal Neurons

Lin Yang[1,3], Zihao Su[1,3], Xiaolei Song[1], Zhenmeiyu Li[1], Ziwu Wang[1], Haotian Chen[1], Zicong Shang[1], Heng Du[1], Yan Wen[1], Guoping Liu[1], Dashi Qi[2], Xiaosu Li[1], Yan You[1], Zhejun Xu[1*] and Zhuangzhi Zhang[1*]

[1]State Key Laboratory of Medical Neurobiology, Institutes of Brain Science, MOE Frontier Research Center for Brain Science, Department of Neurology, Zhongshan Hospital, Fudan University, Shanghai 200032, PR China.

[2]Shanghai Key Lab of Birth Defect, Children's Hospital of Fudan University, Shanghai 200032, PR China.

Corresponding author：

Zhejun Xu, Email: 14111520028@fudan.edu.cn

Zhuangzhi Zhang, Email: zz_zhang@fudan.edu.cn

State Key Laboratory of Medical Neurobiology, Institutes of Brain Science, MOE Frontier Research Center for Brain Science, Department of Neurology, Zhongshan Hospital, Fudan University, 138 Yi Xue Yuan Road, Shanghai 200032, China.

[3]These authors contributed equally to this work.

Abstract

The striatum is structurally highly diverse and its organ functionality critically depends on a normal embryonic development. Although several studies on the gene functional changes that occur during striatal development, a system-wide analysis of the underlying molecular changes is lacking. Here, we present a comprehensive temporal transcriptome atlas of the mouse striatum at multiple developmental stages. Quantitative analysis of 15443 gene products allows us to explore the trajectory of striatal development and identify the correlation between the striatal development and Huntington's disease (HD). More importantly, we provided the global landscape of 277 transcription factor (TF) networks based on co-expression analysis and machine learning. Furthermore, we identified the hub TFs Six3 and Meis2 which are involved in regulating the apoptosis of striatal neurons. Finally, using conditional knockout (CKO) mice and RNA-Seq data, we verified that Six3 and Meis2 indeed regulated neuronal apoptosis and inferred their downstream targets in the striatum.

Introduction

The striatum is the major information processing center in the basal ganglia circuits of the mammalian forebrain. It is involved in multiple neurological functions ranging from motor control, cognition, emotion, and reinforcement to plasticity underlying learning and memory[1]. Cell division, differentiation, migration and apoptosis are essential for striatal organization and function[2-5]. The dysfunction of neurotransmission or neurodegeneration of the striatum results in a number of neurological disorders, including Huntington's disease (HD), Parkinson's disease and obsessive–compulsive disorders[6-8]. Thus, exploring the development of the striatum may provide some new thoughts for disease treatment. The vast majority of neurons (90–95%) in the mouse striatum are GABAergic medium-sized spiny projection neurons (MSNs) which are derived from lateral ganglionic eminence (LGE) progenitors[3,9]. Approximately 5–10% of striatal neurons are aspiny interneurons that are derived from the medial ganglionic eminence (MGE) progenitors[10,11]. The ontogeny of the striatum begins as early as embryonic day 11 (E11.0)[9,12] and then rapidly increases. Subsequently, the striatal interneurons migrate from the MGE into the striatal primordium[11]. Then, cells that originate from the LGE and MGE start to differentiate into different cell types to build on a three-dimensional cellular organization.

During striatal development, transcription factors (TFs) and genetic regulatory networks (GRNs) are essential for the differentiation, migration and survival of the striatal neurons. Currently, most studies have focused on a few genes through specific individual pathways to regulate striatal development. Based on these efforts, the key roles of several transcriptional factors (Gsx2, Dlx1/2, Ascl1, Sp8/9, Isl1, Six3, etc.) in controlling striatal development have been explored[2,5,13-15]. However, many novel TFs or TF functions need to be explored in striatal development. Furthermore, although our knowledge of striatal development has been enhanced by substantial data from reports, their limited focus hampers our understanding of the whole process in the context of systems biology. In fact, TFs are highly dynamic and interactive, and TF networks regulate striatal development though highly coordinated expression changes[16]. To solve these problems, genome-wide transcriptional profiling may shed new light on the

molecular mechanisms of striatal development and help to define novel TF networks involved in striatal development and predict novel functions of hub genes in TF networks.

Here, we aimed to gain novel insight into striatal development based on transcriptional regulation. Gene co-expression network analysis is a main approach to annotate unknown gene networks and putative novel gene functions[17], which has been utilized successfully to provide new biological insight into gene networks involved in several developmental processes, such as fetal brain development[18] and fertilized egg development[19]. Constructing co-expression networks is generally straightforward and here are some handy software packages, such as WGCNA[20]. However, the resulting network of connected genes may become very complex and lack biological significance, which limits its biological interpretation[21]. Therefore, a strategy consistent with biological questions is needed to infer reliable networks. The strategy of using a priori knowledge and data integration can further enhance the elucidation of gene regulatory functions[22]. By using known functional annotations as labels for given genes, one can determine the extent to which the network summarizes the information by using a machine learning model to classify these genes as belonging to a provided function based on their correlation in the co-expression network[23].

In the present work, we analyzed the developmental process of the striatum from a macro-scale perspective and found that HD was closely related to the later stage of striatal development. Moreover, we provided the global landscape of 277 TF networks in striatal development based on a machine learning model for the first time, and these TF networks can be used to infer potential transcription factor functions. For example, our results showed that Six3 and Meis2 were included in the TF networks involved in regulating neuronal apoptosis. Using conditional knockout (CKO) mice, we verified that Six3 and Meis2 indeed regulated the neuronal apoptosis during striatal development. These results further validate the rationality of our global landscape of TF networks. Furthermore, we predicted the downstream targets of Six3 and Meis2 by analyzing the RNA-Seq data of the Six3 and Meis2 CKO mice, respectively. In conclusion, this study provides the mRNA atlas and global TF networks, which offer valuable assets and

provide direct insight into the molecular details of the regulation of the striatal developmental process that enable further data interrogation to obtain clues for a better understanding of striatal development and diseases.

Results

Atlas of striatal development in mice

Identify the novel functions of the TFs in the mouse striatum by the workflow (Fig. 1). After calculating the normalized gene expression level as fragments per kilobase of transcript per million mapped reads (FPKM; see Supplementary Methods), we first performed correlation analysis of the experimental data between 5 time-points. To facilitate the following description, we defined four stages according to our data (from E13.5 to E16.5: stage I; from E16.5 to P0: stage II; from P0 to P14: stage III; from P14 to P60: stage IV). We found that all data showed good consistency and a high degree of correlation (R = 0.95–0.99) between the temporally adjacent experiments (Supplementary Fig. 1a). Then, we performed correlation analysis of experiments between two adjacent time-points (Supplementary Fig. 1b). Our results showed that the Pearson correlation coefficient of stage II is higher than that of other stages, and the Euclidean distance is lower than that of other stages. From the perspective of data structure, the two networks of data (E16.5 and P0) are very similar, thus there will not be excessive changes in gene expression in stage II. After pre-processing the raw data (see Supplementary Methods), a list of 15443 genes was compiled from all 5 time-points that may represent the core components of the striatal transcriptome (Supplementary Fig. 1c; Supplementary Table 1).

To further investigate the developmental status of the striatum at each stage, we obtained the differentially expressed genes (DEGs) at each stage (see Supplementary Methods). We identified all 8861 DEGs, including 882 (10%) TFs in 4 stages (Fig. 2a; Supplementary Table 2), and then performed biological process and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment to assign genes to functional categories for each stage (Supplementary Table 2). The functions of these DEGs were significantly enriched in biological processes, including “regulation of cell cycle” (stage I), “regulation of synapse organization and axonogenesis” (stage II), “regulation of ion transmembrane transport” (stage III), “positive regulation of intrinsic apoptotic signaling pathway and ATP synthesis” (stage IV) and other biological processes (Supplementary Fig. 2a). The functions of these DEGs were significantly

enriched in pathways, including “cell cycle” (stage I), “axon guidance” (stage II), “calcium signaling pathway” (stage III), “Huntington disease” (stage IV) and other pathways (Supplementary Fig. 2b). From the whole process of striatal development, our results indicated that cell division gradually decreased. In contrast, neuronal differentiation, synapse organization and the dopamine signaling pathway were gradually enhanced from stage I to stage III (Fig. 2b). In addition, the ATP synthesis, the oxidative stress response and the intrinsic apoptotic signaling pathway were increased in stage IV (Fig. 2b). Biological processes and pathways are important for the development of the striatum, but when they are disrupted, they can lead to a number of neurodegenerative diseases, such as HD and Parkinson's disease.

Correlation of striatal development and HD

HD is the most familiar striatal degenerative disease[24] and it has been suggested to resemble a misregulated embryonic development process in a variety of ways, and this view is supported by experimental evidence from a limited number of genes[25,26]. Here, we found that the significantly altered genes from P14 to P60 are strongly associated with Huntington's disease in the striatum (Fig. 2c). Therefore, the mRNA atlas of the mouse striatal development dataset obtained in this study and a collected transcriptome dataset of HD obtained from 38 patients[27] enabled us to investigate the correlation of striatal development and HD at the mRNA level. To date, many disease-related dysfunctions have been reported, including dysregulation of the calcium signaling pathway[28], dysregulation of dopaminergic signaling[29] and metabolic abnormalities[30-34]. These events can also lead to downstream changes in gene expression. To correlate striatal development with HD on the levels of pathways, we listed all gene products that were detected in two studies and counted the number of significantly altered gene products in one of the seven pathways known to be involved in development and HD (Fig. 2d, e; Supplementary Table 3). Two-tailed hypergeometric statistical tests (see Supplementary Methods) were carried out to determine whether there was a significant overlap of altered gene products on a specific pathway between striatal development and HD. Importantly, all pathways

exhibited high correlations ($-\log_{10}$P-value > 1.3). Hence, these analyses provided evidence that striatal development was closely associated with HD processes.

To obtain a global view of when the elevated genes identified in HD are expressed in striatal development, we selected 57 genes that were differentially expressed in the striatum of HD patients and then mapped the expression of their mouse homologs to the 5 time-points during striatal development (Fig. 2f). A considerably higher percentage of genes was identified in the postnatal days (P0-P60) than in the earlier time-points (Fig. 2f). Altogether, with the previously described findings, our results showed that HD shared many characteristics with striatal development and suggested that the disease may result from the dysregulation of pathways.

Co-expression analysis identified inter-related functional modules of continuously expressed genes

To further provide insights into the functional transitions during striatal development, we divided the sum of DEGs obtained into periodically expressed genes (PEGs) and continuously expressed genes (CEGs) (see Supplementary Methods; Supplementary Table 4). We used all 1257 PEGs, including 115 (9%) TFs in 4 stages, and then performed biological process enrichment to assign genes to functional categories for each stage (Supplementary Fig. 3a, b). Our results show that the PEG biological process is in line with our delineative striatal development model in Figure 2. We applied weighted gene co-expression network analysis (WGCNA)[20] to obtain a system-wide understanding of groups of CEGs (7095 genes, including 740 TFs) (Supplementary Table 5) whose co-expression patterns are highly correlated during striatal development. The soft-threshold power was defined as 13, the scale-free topology index was > 0.85, and it was thus used in this study (Supplementary Fig. 4a). The hierarchical clustering dendrogram represents co-expressed genes that were highly correlated, and the relative expression levels of these genes were represented (Fig. 3a). In total, 6 distinct co-expression modules were identified that showed a diverse trend of gene expression, and each module was assigned a colour name (Fig. 3a; Supplementary Table 5). Then, to show trends in gene expression, the genes from

each module were grouped into two trend clusters using the fuzzy c-means algorithm[35] (Fig. 3a). Clearly, each module contains not only positively correlated genes but also negatively correlated genes. In addition, the network heat-map plot further suggested those main modules were independent from each other (Supplementary Fig. 4b).

In addition, biological process and KEGG enrichment analysis of the gene members in each module revealed distinct biological functions and pathways (Supplementary Fig. 5a, b; Supplementary Table 5). These findings not only make it possible to understand the dynamic events during striatal development, but also may help identify key factors. Intriguingly, genes in the orange module were keeping going up or keeping going down all the time in 5 time-points (Fig. 3a). Genes in this module are enriched in “mitotic nuclear division” and “DNA replication” (Supplementary Fig. 5a, b), consistent with cell division and neuronal differentiation during all stages. Many important genes are essential for striatal MSN proliferation and differentiation, including Gsx2, Dlx1/2 and Sp9, which were defined in this module (Fig. 3b). This was consistent with previous reports that Gsx2, Dlx1/2, and Sp9 were closely related to cell proliferation and differentiation in striatal development[3,4,13]. For example, Sp9 is highly enriched in D2-type MSNs and marks striatal precursors and MSNs from their earliest stage into adulthood. In Sp9 mutant mice, nearly all D2-type MSNs are lost[4]. Thus, Sp9 is required for the differentiation of D2-type MSNs. In contrast, genes in the red module were maintained up until stage II, and sharply down in stage III (Fig. 3a). Genes in this module were enriched in pathways related to “synapse organization”, “regulation of transmembrane transporter activity”, and “neurotransmitter transport”, all related to synaptic transmission (Supplementary Fig. 5a). Some genes were essential for striatal development, including Zfhx3, Adora2a, Ppp1r1b and Drd2, which were defined in this module (Fig. 3b). These genes were crucial for normal development and functions of striatal MSNs. In previous studies, Zfhx3, Adora2a, Ppp1r1b and Drd2 were closely related to MSN differentiation and survival [4,36-38]. Thus, our co-expression modules help us to identify the key genes involved in regulating the development of MSNs. Other pathways in the blue module include “oxidative phosphorylation”, “citrate cycle”, and “cell cycle” (Supplementary Fig. 5a, b), which are necessary to produce energy in

support of embryonic growth and cell division. To further confirm that the co-expression module was reasonable, our data were solid. We selected genes randomly and verified the expression trend of these genes (Zfhx3, Penk, Drd1, Tac1 and Isl1) by in situ hybridization (Fig. 3c). Our results showed that the expression trend of these genes was corrected with our co-expression modules. Altogether, these results not only make it possible to understand the dynamic events during striatal development, but also may help to identify key factors.

Prediction of potential functional TF networks based on machine learning

Gene co-expression networks can be used to associate genes of unknown function with biological processes, to prioritize candidate disease genes or to discern transcriptional regulatory programs[17,23]. However, common co-expression analyses are based entirely on correlations rather than biological significance, making them often less than satisfactory[21]. To make up for this, we combined co-expression analysis with machine learning to predict a number of potentially significant gene networks. From our previous analysis, we identified transcriptional regulators that exhibited clear temporal and lineage dynamics. This prompted us to focus on TF expression in our time series. We identified 740 differentially expressed regulators from six CEGs co-expression modules (Fig. 4a; Supplementary Table 5) and these TFs were predicted by functional TF networks based on a machine learning model.

Using the co-expression of TFs with optimized parameter settings and known functional sets (Supplementary Table 6), machine learning was performed by following the following steps (see Supplementary Methods): calculate correlation metrics (five algorithms), machine learning predicts interactions, clustering high-confidence interactions into functional networks (Supplementary Fig. 6a). The output of machine learning was well organized and ready for subsequent analysis to predict functional TF networks (Fig. 4b). The distribution of the interaction network node degree distribution is a power-law (Supplementary Fig. 6b), which has certain biological significance and this phenomenon is also widely confirmed[39]. The high proportion of human homologous TFs[40] and nervous system disease TFs[41] in the mice involved in these

networks suggests that the subsequent analysis is likely to provide a reference for human striatum studies (Supplementary Fig. 6c, d). After that, we parted the network using the ClusterONE algorithm to part TF networks for downstream analysis[42]. To assess the physiological significance of the putative striatum functional networks, we analyzed the TF networks for coherent biological functions based on Gene Ontology (GO) annotations. What is exciting is that, almost all of the TF networks in striatal development were significantly enriched for associations with essential processes (Supplementary Table 7). To prove that the predicted TF networks obtained through machine learning have more biological significance than just co-expression analysis, we used the method of comparison with the random model to assess the accuracy of the predicted TF networks (see Supplementary Methods). For each module, measured as the biological process enrichment adjusted p-value distribution, there was markedly higher significance for the functional of putative TF networks compared to the same number of randomized networks of TFs in the network that were not predicted (Fig. 4c, Supplementary Fig. 7). Therefore, 277 putative TF networks were identified from five co-expression modules (Fig. 4d; Supplementary Table 7).

After that, we roughly predicted some novel TF functions. For example, we predicted that TF network 1 was enriched in “Regulation of neuron apoptotic process” with high significance (adjusted P 5.79E-06) (Fig. 4e). In this TF network, these TFs (Grin1, Egr1, Ppargc1a, Lig4, Msh2, Nono and Npm1) have been reported to be associated with neuronal apoptotic processes[43-46]. For instance, Grin1 mutations cause neurodegeneration in the striatum and increase neuronal death in early development[43]. Likewise, we predicted that TF network 2 was enriched to “Central nervous system neuron differentiation” with high significance (adjusted P 4.97E-13) (Fig. 4e). In this TF network, these TFs (Ascl1, Gsx2, Dlx1, Lhx8, Sox1, Sufu, Arx, Sall1 and Sall3) have been reported to be associated with cell differentiation[2,3,5,47-53]. Indeed, Ascl1, Gsx2 and Dlx1 were highly expressed in the LGE during the development of the striatum. Deletion of the expression of these genes leads to abnormal differentiation of striatal MSNs[2,3,5]. We also predicted other functional TF networks enriched in “Cell fate commitment”, “Axonogenesis”, and “Cell-cell signaling by wnt” (Fig. 4e). These

functional TF networks during striatal development can provide us with a rich treasure house to predict novel functional genes, which will be helpful for researchers to carry out work related to the striatum, including exploring the core transcriptional regulators, constructing transcriptional regulatory networks and even designing drugs related to disease.

Inference of novel functional hub TFs from putative TF networks

The regulation of TF is important in the process of striatal development, but the function of most TFs remains to be further explored. Some methods can realize functional annotation of nodes and modules and reveal the essence of gene function[23]. Enrichment analysis for the genes within a module is the most widely used strategy for associating modules with particular functions. Under the “guilt-by-association” rule, these functional modules provide a powerful framework for identifying novel genes related to biological processes and their functional annotations in the absence of strong a priori knowledge[22]. For example, the TF network mentioned above (Fig. 4e), is significantly enriched for biological processes including three main categories “neuron apoptotic process”, “circadian rhythm” and “learning or memory” (Fig. 5a) based on the GO term relationships using BiNGO[54]. The “circadian rhythm” and “learning or memory” suggest that the striatum may be involved in some behavioral functions, such as sleep/wake cycles and spatial memory. Then, we used the TFs of appropriate GO terms with clear functional bias (regulation of neuron apoptotic process/ adjusted P: 3.53E-06, circadian rhythm/ adjusted P: 2.02E-22, learning or memory/ adjusted P: 9.66E-06) (Supplementary Table 7) as reference standards for downstream functional gene exploration. To test the biological relevance of these networks in vivo, we reconstructed the network structure of TFs within each of these biological processes based solely on the Algorithm for the Reconstruction of Accurate Cellular Networks (ARACNE)[55] to explore novel functional hub genes (see Supplementary Methods). Hub genes are highly connected genes within a module that can be predicted to regulate the expression of many other module memberships[56]. We reconstructed the potential transcription factors regulatory networks for each biological process with a

number of core factors that have the most connectivity. In addition to the known functional reference TFs (grey nodes) (Fig. 5b; Supplementary Fig. 8a and b), we infer some hub TFs based on the connectivity of network nodes, which are more likely than others to be related to the function. For example, we hypothesize that Zfhx3, Six3 and Meis2 are most likely involved in “regulation of neuron apoptotic process” (Fig. 5b). Likewise, in the “circadian rhythm”, we predicted that Camk4, Msh2 and Lig4 are at the core in the network (Supplementary Fig. 8a). We also inferred that Zfhx3, Nono and Msh2 are related to “learning or memory” (Supplementary Fig. 8b).

CKO of putative TFs leads to abnormal apoptosis in striatal neurons

Neuron apoptosis is very important in the development of the nervous system, and the striatum is no exception. Abnormal neuronal apoptosis leads to many neurological diseases, such as HD and Parkinson's disease, which are closely related to the striatum. Therefore, it is necessary to explore neuronal apoptosis during striatal development, which will provide new insights and ideas for researchers. Next, we mainly verified the predicted neuronal apoptosis-related TFs to further confirm that our global landscape of TF networks was reliable. To test this hypothesis, we selected several core factors: Zfhx3, Six3 and Meis2 (Fig. 5b), in the center of the network, all of which were poorly studied in “regulation of neuron apoptotic process” in the striatum. Recently, the TF Zfhx3 was reported to induce abnormal cell apoptosis in the striatum during the first postnatal week. However, whether Six3 and Meis2 are involved in the cell apoptosis of striatal MSNs is largely unknown.

To investigate whether Six3 plays a functional role in striatal MSN apoptosis, we conditionally deleted the Six3 gene in the LGE using Dlx5/6-Cre lines (henceforth described as Six3-CKO). Cre recombinase was robustly expressed in the LGE and Six3 was removed from the striatum. Next, we performed immunostaining of cleaved Caspase-3 at P0. Our results showed that the cleaved Caspase-3 positive cells were increased approximately 2-fold in the Six3-CKO mice compared with their littermate controls (Fig. 5c, d). To investigate the potential mechanism by which Six3 affects neuronal apoptosis in the striatum, we performed RNA-Seq analysis at P0 between

Six3-CKO mice and their littermate controls. Our RNA-Seq data revealed that approximately 431 DEGs in the Six3-CKO mice (220 downregulated and 211 upregulated RNA expression) and then performed biological process enrichment of these DEGs to functional categories (Fig. 5e, Supplementary Table 8), suggesting that Six3 plays a key role in regulating neuronal apoptosis of the LGE at P0 (regulation of neuron apoptotic process/ adjusted P: 5.45E-05/ 16 DEGs). To further explore the potential hub downstream factors for regulating neuronal apoptosis of Six3, we constructed the GRN of 16 significantly changed genes using Genie3 (see Supplementary Methods), which predicted the directed regulatory links from genes to genes (Fig. 5f). To determine which genes might be controlled by Six3 in the striatum, we also measured the static network parameter of this GRN. Adora2a was thought to have high pleiotropic effects and play essential roles in regulating neuronal apoptosis (Fig. 5g). Changes in Adora2a were expected to have larger biological impacts than those occurring in genes in the network periphery. This suggests that Adora2a may be a potential hub downstream factor for Six3 to regulate neuronal apoptosis. Adora2a is highly expressed in D2-type MSNs and is involved in cell apoptosis[4,57]. These findings suggest that Six3 regulated striatal MSN apoptosis by regulating the expression of Adora2a.

To further verify the rationality of our global landscape of TF networks and confirm whether Meis2 was involved in striatal MSN apoptosis, we designed the Meis2 targeting vector to generate the Meis2^{Flox} allele through the strategy of homologous recombination (Fig. 6a). To analyze whether Meis2 plays a functional role in MSN apoptosis, we used Dlx5/6-Cre to conditionally knockout Meis2 (henceforth described as Meis2-CKO) in the striatum. Our results showed that Meis2 mRNA expression was complete in the LGE or striatum in Meis2-CKO mice compared with that in their controls by in situ hybridization (Fig. 6b). Note that the expression of Meis2 in the striatum was complete but the expression of Meis2 in the cortex remained. Next, we performed immunostaining of cleaved Caspase-3 at P0. There were more cleaved Caspase-3 positive cells in the Meis2-CKO mice than in their controls (Fig. 6c, d). In order to investigate the potential mechanism of Meis2 affecting neuronal apoptosis in

the LGE or striatum, we performed RNA-Seq analysis at E16.5 between Meis2-CKO mice and their littermate controls. Our RNA-Seq revealed that approximately 858 DEGs in Meis2-CKO mice (632 downregulated and 226 upregulated RNA expression) and then performed biological process enrichment of these DEGs to functional categories (Fig. 6e; Supplementary Table 9), suggesting that Meis2 plays a key role in neuron apoptosis of the striatum (regulation of neuron apoptotic process/ adjusted P: 9.65E-05/ 22 DEGs). To further explore the potential hub downstream factors for regulating neuronal apoptosis of Meis2, we constructed the GRN of 22 significantly changed genes using Genie3 (see Supplementary Methods), which predicted the directed regulatory links from genes to genes (Fig. 6f). To determine which genes might be controlled by Meis2 in the striatum, we also measured the static network parameter of this GRN. Isl1 and Adora2a were thought to have high pleiotropic effects and play essential roles in regulating neuronal apoptosis (Fig. 6g). Changes in Isl1 or Adora2a were expected to have larger biological impacts than those occurring in genes in the network periphery. Our results suggest that Isl1 and Adora2a may be potential downstream factors for Meis2 to regulate the neuronal apoptosis. In fact, Isl1 and Adora2a have been reported to be expressed in D1-type and D2-type MSNs and regulate cell apoptosis in the striatum[4,14,57]. Altogether, these results indicated that the transcriptional factors Six3 and Meis2 are the hub genes in the transcriptional network for the survival of striatal MSNs.

Discussion

The current study presented the first mRNA panoramic overview of the developing mouse striatum from the embryonic stage to adulthood. More critically, we provided the global landscape of TF networks in striatal development for the first time based on machine learning. Based on the TF networks, we identified two key factors (Six3 and Meis2), that are critical for the survival of striatal MSNs. Taking advantage of the Six3 and Meis2 mutant RNA-Seq data, we inferred their direct targets in regulating the survival of the MSNs. This mRNA atlas and global TF networks provide a systematic view for striatal development and permit data mining and knowledge generation as a public resource for other scientists.

During mouse striatal development, our preliminary bioinformatic analysis suggested that cell division was gradually decreased, but the neuronal differentiation, synapse organization and the dopamine signaling pathway were gradually increased. It is worth mentioning that the DEGs in the striatum from P14 to P60 are closely related to HD. This phenomenon inspired us to investigate whether there is experimental evidence for a connection between striatal development and HD. By comparing the mRNA expression of striatal development with that of HD[27], we found that a noticeable number of pathway regulators in HD are also differentially expressed during striatal development. Moreover, we identified statistically significant networks of overlapping genes in seven pathways between striatal development and HD, which suggests that HD regains some features of striatum embryonic development but loses some basic functions of the adult striatum and that HD is likely a consequence of the dysregulation of master regulators and pathways.

The main motivation of this study is the identification of networks of co-expressed TFs, with inherent expression distribution corresponding to discrete biological processes to be associated with striatal development. In order to make up for co-expression alone deficiency, we adopt a new strategy to optimize the co-expression analysis. Individual pair-wise TF associations were scored based on the temporal expression profile similarity measured during the striatal development. Next, we used an integrative computational scoring procedure to derive TF interactions[58]. The support

vector machine learning classifier used was trained on correlation scores obtained for reference annotated TF sets[58], and combined all of the input co-expression data together with previously published TF interactions evidence[59]. Measurements of overall performance showed high area under the curve by the TF co-expression data. The final filtered 277 TF networks consists of 655 transcription factors. From these TF networks, we identified transcriptional regulators across all networks, including Ascl1 Gsx2 and Dlx1 were highly co-expressed in the LGE during the development of the striatum, along with their co-expressed targets. Deletion the expression of these gene leads to abnormal differentiation of striatal MSNs[2,3,5]. Thus, these TF networks can be used to speculate novel gene functions during striatal development. For example, we identified two key factors (Six3 and Meis2) from one TF network related to neuronal apoptosis. Since neuronal apoptosis is one of the major events in striatal development, we further explored the possible downstream factors of these novel functional genes to elucidate their potential mechanisms underlying neuronal apoptosis. Using Six3-CKO mice by performing RNA-Seq, we speculated that Six3 regulated survival of the striatal neurons by regulating the expression of Adora2a. Adora2a is highly expressed in D2-type MSNs and is involved in cell apoptosis[4,57]. Moreover, we generated Meis2-CKO mice for the first time. Using the same approach, we predicted that Isl1 and Adora2a were regulated by Meis2 in regulating the neuronal apoptosis. These findings were supported by previous reports[4,14,57]. This raises the exciting potential to identify more critical factors from our data and deepen our understanding of the regulation of striatal development.

In conclusion, the mRNA atlas and global TF networks provide valuable assets and provide direct insight into the molecular details of the regulation of the developmental process that enable further data interrogation to obtain clues for a better understanding of striatal development.

Methods

Animals and tissue collection.

The mouse genetic background was CD1 in this study. The Dlx5/6-Cre and Six3$^{F/F}$ mice were previously described [47,60]. We generated Meis2$^{F/+}$ mice in this study. Details of the targeting construct are shown in Figure 6. Briefly, we first generated Meis2 knockout mice via the ESC targeting system. The cassette is composed of an FRT site followed by a LacZ sequence and a loxP site. This first loxP site is followed by neomycin under the control of the human beta-actin promoter, SV40 polyA, a second FRT site and a second loxP site. A third loxP site is inserted downstream of the targeted exon 8. Exon 8 is thus flanked by the loxP site. A “conditional ready” (floxed) allele can be created by flp (ACT-flp) recombinase expression in mice carrying this allele. Subsequent Cre expression results in knockout mice. The Mesi2$^{F/F}$ mice genotype was identified by PCR (Polymerase Chain Reaction). The primer: Meis2-loxp F1: TGTCAAACATTGCCATCCAAACAGT; Meis2-loxp F2 TCATGGAGAAGACAGCGCTGGT. Mice were group-housed with approximately three mice in a cage on a 12-h light, 12-h dark cycle, with ad libitum chow and water. The day of vaginal plug detection was calculated as E0.5, and the day of birth was considered as P0. Males and females were both used in all experiments. All animal experiments described in this study were approved in accordance with institutional guidelines at Fudan University Shanghai Medical College.

RNA-Seq analysis was performed as previously described[61]. For the temporal transcriptome analysis, the whole LGE or striatum was separated from embryonic to adult mice. LGE or striatum were collected at 5 stages covering the embryonic and postnatal stages: E13.5, E16.5, P0, P14 and P60. Each stage had three biological replicates, but we obtained only two successful biological replicates at P14. The whole LGE or striatum tissues were washed twice with ice-cold HBSS (Hanks’ Balanced Salt Solution) to remove blood and quick-frozen in liquid nitrogen and then stored at –80 °C for total RNA extractions. The striatum (including VZ, SVZ and MZ-striatum) of the Dlx5/6-Cre; Six3$^{F/F}$ (henceforth described as Six3-CKO) mice and their Dlx5/6-Cre littermates (henceforth described as controls) were dissected at P0 (n=3 mice, each

group). The LGE or striatum (including VZ, SVZ and MZ-striatum) of the Dlx5/6-Cre; Meis2$^{F/F}$ (henceforth described as Meis2-CKO) and their Dlx5/6-Cre littermates (henceforth described as controls) were dissected at E16.5 (n=3 mice, each group). Data from this experiment have been deposited in the GEO database (GSEXXXXXX).

Immunohistochemistry.

Mice were anaesthetized with an intraperitoneal injection of ketamine (100 mg per kg) and transcardially perfused with a fixative solution containing 4% paraformaldehyde (PFA). Brains were post-fixed in 4% PFA at 4 °C overnight. Sections of 20-μm thickness were used for immunostaining. Sections were washed with 0.05M TBS for 10 min, incubated in Triton-X-100 (0.5% in 0.05 M TBS) for 30 min at room temperature (RT), and then incubated with block solution (10% donkey serum +0.5% Triton-X-100 in 0.05 M TBS, pH=7.2) for 2 h at RT. Primary antibodies against cleaved Caspase-3 (1:500; Cell Signaling, 9661, 43) were diluted in 10% donkey serum block solution and incubated overnight at 4 °C, then rinsed 3 times with 0.05M TBS. Secondary antibodies (from Jackson, 1:500) matching the appropriate species were incubated for 3 h at RT. Fluorescently stained sections were then washed 3 times with 0.05 M TBS. This was followed by 4′, 6-diamidino-2-phenylindole (DAPI) (Sigma, 200 ng/ml) staining for 3 min and sections were then cover-slipped with Gel/ Mount (Biomeda, Foster City, CA). Fluorescence images were collected with an Olympus-BX61VS microscope using 10× and 20× objectives.

In *situ* RNA hybridization

In *situ* RNA hybridization experiments were performed using digoxigenin labelled riboprobes on 20-μm frozen sections as previously described[13,36] or made from cDNAs amplified by PCR using the following primers:

Drd1 Fwd: GGCATTTGGAGAGATGTGGCACCAG.

Drd1 Rev: AGATAGCCCAATACCTGTCCACGCT.

Meis2 Fwd: CGATGGGTTAGACAACAGCGTAGCTTCACCTG.

Meis2 Rev: ATGGCTTGGCAAATATGATGCATTGGGTCCATG.

Isl1 Fwd: ATGGGAGACATGGGCGATCC.

Isl1 Rev: CATGCCTCAATAGGACTGGCTACC.

Penk Fwd: TCGGAAGGACAGGATGTCATCA.

Penk Rev: CGTCAGGAGAGATGAGGTAACAAAC.

Tac1 Fwd: CCCCTGAACGCACTATCTATTC.

Tac1 Rev: TAGAGTCAAATACCGAAGTCTCAG.

Zfhx3 Fwd: CGATCTGGCCCAGCTCTACCA.

Zfhx3 Rev: CTGTAAGCCTGCGAGGGCATAG.

RNA sequencing and data processing.

Total RNA was extracted from the LGE or striatum tissues and treated with deoxyribonuclease I (DNase I). mRNA was isolated by Oligo Magnetic Beads and cut into small fragments that served as templates for cDNA synthesis. Once short cDNA fragments were purified, they were extended with single nucleotide adenines, ligated with suitable adapters, and amplified by PCR before they were sequenced. High-throughput RNA sequencing (RNA-Seq) experiments were carried out using the Illumina comprehensive next-generation sequencing (NGS) technique. Raw data were filtered and processed by the FastQC software (Version 0.11.5, Available online at website: http://www.bioinformatics. babraham.ac.uk/projects/fastqc). Low quality RNA-Seq reads were removed if they had a Phred quality score less than 20 or had less than 50 nucleotides. The filtered reads were mapped onto the mouse reference genome (GRCm38.p2.genome, released on 12/10/2013) using HISAT2 software (Version 2.1.0)64. Assembly and quantification of the transcripts were accomplished with StringTie software (Version 1.3.1) using the mouse genome annotation file as the reference (gencode.vM2. annotation, available at the GENCODE website) 65. The number of FPKM was used for the measurements of the relative abundances of the transcripts.

The Pearson correlation coefficient was calculated between biological replicates with the normalized expression levels of $\log_2$ (FPKM value +1). The missing value was default to 0. Comparison of the two biological replicates showed that the expression

values between them were highly correlated (average R = 0.98). Hence, the average FPKM value of the two replicates was taken as the expression level for the sample at each time-point. However, before that, the transcription was at least twice in biological repetition. To reduce the influence of transcription noise, a gene was defined as expressed if its FPKM value was ≥1 at least at a time-point.

The read counts were utilized for differential expression analyses using the edgeR package in R. A key feature of the edgeR package is the use of weighted likelihood methods to implement a flexible empirical Bayes approach in the absence of easily tractable sampling distributions[62]. The false discovery rate (FDR) was controlled by using the Benjamini and Hochberg (BH) algorithm. Unless otherwise specified, genes with FDR < 0.01 and fold-change (FC) > 2 or < 0.5 were judged to be differentially expressed.

Hierarchical clustering was performed by the pheatmap package in R. Additionally, the transformed and normalized gene expression values were calculated by dividing their expression level at different time-points with their maximum observed FPKM for hierarchical clustering.

Referenced datasets

We obtained microarray gene expression profiling data from the following GEO Datasets (http://www.ncbi.nlm.nih.gov/geo/): GSE3790 which includes human post-mortem striatal tissue from 38 HD cases and 32 controls[27]. For the Affymetrix microarrays, we downloaded the gene expression matrix to identify the DEGs (adjusted P < 0.05) by using the limma package.

Functional annotation GO term analysis

Functional annotation for the expressed gene list was performed with THE clusterProfiler, GO.db, DOSE, org.Mm.eg.db and org.Hs.eg.db packages in R. The adjusted p-value method was controlled by using the BH algorithm. GO terms with adjusted p-values < 0.05 were judged to be significant terms. The enrichGO function was used to make GO enrichment and the enrichKEGG function was used to make

KEGG enrichment.

The network diagram of GO term relationships is generated by BiNGO (Biological Networks Gene Ontology tool) which is an open-source Java tool to determine which GO terms are significantly overrepresented in a network of genes[54].

Two-tailed hypergeometric statistical test

Two-tailed hypergeometric statistical tests were carried out by R to determine whether there was a significant overlap (p-value < 0.05) of altered genes on a specific pathway between striatal development and HD.

Weighted gene co-expression network analysis

Weighted gene co-expression network analysis was performed using R package WGCNA [20]. The $\log_2$ (FPKM value +1) transformed data were used to generate a matrix of the Pearson correlations between all pairs of genes across the samples. The value $\beta = 13$ was chosen as a saturation level for a soft threshold of the correlation matrix, based on the criterion of approximate scale-free topology. To minimize effects of noise and spurious connections, the correlation matrix was transformed to the topological overlap matrix using a “TOMsimilarity” function implemented[20]. The topological overlap matrix was then used to group highly co-expressed genes by performing average linkage hierarchical clustering. The dynamic hybrid tree cut algorithm was used to cut the hierarchal clustering tree[20], and modules were defined as the branches resulting from this tree cutting. The minimum module size was defined as 30 and the threshold for the merging of similar modules was defined as 0.35.

Then, in order to show trends in gene expression, the genes from each module were grouped into clusters using the Mfuzz package in R with the fuzzy c-means algorithm[35]. Before clustering, standardisation of the expression values of every gene was carried out, so that the average expression value for each gene was 0 and the standardization deviation of its expression profile was 1.

Processing of machine learning

Since associated functions within a TF network are expected to share co-expression patterns, the pairwise TF profile similarity was computed using up to five correlation algorithms (mutual information (MI), Bayes correlation (Bayes), Jaccard similarity coefficient (Jaccard), Pearson correlation coefficient (PCC) and Apex) that emphasize multiple profile features. To increase reliability and reduce computational time, co-expression correlation scores less than 0.75/0.85 were removed. Then, multiple correlation vectors were combined. After that, the combined correlation vectors were input into a supervised machine-learning model that was both trained to predict new TF pairs and benchmarked against reference positive (annotated) TF groupings (that is, enrichment of each module TFs in the GO database) (Supplementary Table 10) and negatives (that is, combinations of TFs in distinct enrichment of each module TF). The machine-learning model was from EPIC (Elution Profile-based Inference of Complexes), which can predict co-expression protein complexes based on known complexes[58].This principle based on this co-expression correlation model can be used to predict functional TF networks by random forest classifiers in Python. The additional supporting evidence was integrated (for example functional interactions inferred from co-expression) from public sources TRRUST2[59], thereby producing richer and more accurate interaction networks. Finally, network-partitioning to define TF network membership by ClusterONE was applied[42]. Moreover, each cluster is compared to annotated TF networks curated in GO. In order to prove that the co-expressed TF networks obtained through machine learning have more biological significance, the functional significance of the randomly generated TF networks was compared. The randomly generated TF network is the random sampling of the TFs involved in machine learning, and the random sampling ensures that the number of TF networks and the total number of genes in each TF network are consistent.

Static network analysis and gene regulatory network prediction

The reconstructed network was imported into the Cytoscape[63] software package for subsequent analyses. The network was analyzed based on the topological parameters (i.e., number of nodes, diameter, radius, centralization, density, heterogeneity, number

of connected components, number of the shortest paths, characteristic path length, and average number of neighbors) and central parameters (i.e., node degree distribution and neighborhood connectivity distribution) using the Network Analyzer plug-in.

To infer the gene regulatory network, given the expression profiles and the list of genes, putative regulatory links from genes to all other genes were estimated using Genie3[64] with default parameters, and only the top 25% of regulatory links with a prediction weight were retained for constructing a GRN. For each node (gene) in the network, in- and out-degree represented the number of edges (regulatory links) directed to and from this node, respectively.

Annotation of genes using existing datasets

The transcription factor annotation integrates three databases (Supplementary Table 10), which are TRRUST2, TFdb and GO. TRRUST2 (https://www.grnpedia.org/trrust/) is a manually curated database of human and mouse transcriptional regulatory networks[59]. TFdb (http://genome.gsc.riken.jp/TFdb/) is a database containing mouse transcription factor genes and their related genes. Genes annotated in the Gene Ontology database as GO:0003700, “DNA binding TF activity”.

Disease annotation comes from the DisGeNET (https://www.disgenet.org/) database which is a discovery platform containing one of the largest publicly available collections of genes and variants associated with human diseases[41]. The genes annotated in the DisGeNET database as Nervous System Diseases, Mental Disorders, Behavior and Behavior Mechanism were collected as custom nervous system associated disease databases (Supplementary Table 10).

Homologous annotation comes from the Brain Atlas (https://www.proteinatlas.org/humanproteome/brain) which explores protein expression in the mammalian brain by visualization and integration of data from three mammalian species (human, pig and mouse).[40]

Acknowledgements

We are grateful to Guillermo Oliver at Northwestern University for the generous gift of Six3$^{F/F}$ mice, and Kenneth Campbell at University of Cincinnati College of Medicine for the Dlx5/6-Cre mice.

Author Contributions

L.Y. performed experiments and analysis and wrote the paper. S.Z. performed experiments and analysis. S.H., Z.L., Z.W., H.C., Z.S., H.D, Y.W, G.L., D.Q., X.L., Z.Y. and Y.Y. helped conduct experiments. Z.X. designed the experiments and analyzed the results. Z.Z. designed the experiments and analyzed the results and wrote the paper.

Potential conflicts of interest

None of the author has any conflict of interest to declare.

Funding

Research grants to Z.Yang. from National Key Research and Development Program of China (2018YFA0108000), National Natural Science Foundation of China (NSFC 31820103006, 31630032, 31425011, and 31429002), research grant to Y. You (NSFC 31700889), research grant to D. Qi (NSFC 81974175).

Data availability statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

References

1 Graybiel, A. M. Building action repertoires: memory and learning functions of the basal ganglia. *Curr Opin Neurobiol* **5**, 733-741, doi:10.1016/0959-4388(95)80100-6 (1995).

2 Waclaw, R. R., Wang, B., Pei, Z., Ehrman, L. A. & Campbell, K. Distinct temporal requirements for the homeobox gene Gsx2 in specifying striatal and olfactory bulb neuronal fates. *Neuron* **63**, 451-465, doi:10.1016/j.neuron.2009.07.015 (2009).

3 Anderson, S. A. *et al.* Mutations of the homeobox genes Dlx-1 and Dlx-2 disrupt the striatal subventricular zone and differentiation of late born striatal neurons. *Neuron* **19**, 27-37, doi:10.1016/s0896-6273(00)80345-1 (1997).

4 Zhang, Q. *et al.* The Zinc Finger Transcription Factor Sp9 Is Required for the Development of Striatopallidal Projection Neurons. *Cell Rep* **16**, 1431-1444, doi:10.1016/j.celrep.2016.06.090 (2016).

5 Castro, D. S. *et al.* A novel function of the proneural factor Ascl1 in progenitor proliferation identified by genome-wide characterization of its targets. *Genes Dev* **25**, 930-945, doi:10.1101/gad.627811 (2011).

6 Wichmann, T. & DeLong, M. R. Functional and pathophysiological models of the basal ganglia. *Curr Opin Neurobiol* **6**, 751-758, doi:10.1016/s0959-4388(96)80024-9 (1996).

7 DeLong, M. R. Primate models of movement disorders of basal ganglia origin. *Trends Neurosci* **13**, 281-285 (1990).

8 Gerfen, C. R. & Surmeier, D. J. Modulation of striatal projection systems by dopamine. *Annu Rev Neurosci* **34**, 441-466, doi:10.1146/annurev-neuro-061010-113641 (2011).

9 Kelly, S. M. *et al.* Radial Glial Lineage Progression and Differential Intermediate Progenitor Amplification Underlie Striatal Compartments and Circuit Organization. *Neuron* **99**, 345-361 e344, doi:10.1016/j.neuron.2018.06.021 (2018).

10 Wang, C. *et al.* Human and monkey striatal interneurons are derived from the medial ganglionic eminence but not from the adult subventricular zone. *J Neurosci* **34**, 10906-10923, doi:10.1523/JNEUROSCI.1758-14.2014 (2014).

11 Marin, O., Anderson, S. A. & Rubenstein, J. L. Origin and molecular specification of striatal interneurons. *J Neurosci* **20**, 6063-6076 (2000).

12 Sheth, A. N. & Bhide, P. G. Concurrent cellular output from two proliferative populations in the early embryonic mouse corpus striatum. *J Comp Neurol* **383**, 220-230 (1997).

13 Xu, Z. *et al.* Sp8 and Sp9 coordinately promote D2-type medium spiny neuron production by activating Six3 expression. *Development*, doi:10.1242/dev.165456 (2018).

14 Lu, K. M., Evans, S. M., Hirano, S. & Liu, F. C. Dual role for Islet-1 in promoting striatonigral and repressing striatopallidal genetic programs to specify striatonigral cell identity. *Proc Natl Acad Sci U S A* **111**, E168-177, doi:10.1073/pnas.1319138111 (2014).

15 Ehrman, L. A. *et al.* The LIM homeobox gene Isl1 is required for the correct development of the striatonigral pathway in the mouse. *Proc Natl Acad Sci U S A* **110**, E4026-4035, doi:10.1073/pnas.1308275110 (2013).

16 Nord, A. S., Pattabiraman, K., Visel, A. & Rubenstein, J. L. Genomic perspectives of transcriptional regulation in forebrain development. *Neuron* **85**, 27-47, doi:10.1016/j.neuron.2014.11.011 (2015).

17 Wu, L. F. *et al.* Large-scale prediction of Saccharomyces cerevisiae gene function using overlapping transcriptional clusters. *Nat Genet* **31**, 255-265, doi:10.1038/ng906 (2002).

18 Spiers, H. *et al.* Methylomic trajectories across human fetal brain development. *Genome Res* **25**, 338-352, doi:10.1101/gr.180273.114 (2015).

19 Gao, Y. *et al.* Protein Expression Landscape of Mouse Embryos during Pre-implantation Development. *Cell Rep* **21**, 3957-3969, doi:10.1016/j.celrep.2017.11.111 (2017).

20 Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 559, doi:10.1186/1471-2105-9-559 (2008).

21 Usadel, B. *et al.* Co-expression tools for plant biology: opportunities for hypothesis generation and caveats. *Plant Cell Environ* **32**, 1633-1651, doi:10.1111/j.1365-3040.2009.02040.x (2009).

22 Wolfe, C. J., Kohane, I. S. & Butte, A. J. Systematic survey reveals general applicability of "guilt-by-association" within gene coexpression networks. *BMC Bioinformatics* **6**, 227, doi:10.1186/1471-2105-6-227 (2005).

23 Ballouz, S., Verleyen, W. & Gillis, J. Guidance for RNA-seq co-expression network construction and analysis: safety in numbers. *Bioinformatics* **31**, 2123-2130, doi:10.1093/bioinformatics/btv118 (2015).

24 Walker, F. O. Huntington's Disease. *Semin Neurol* **27**, 143-150, doi:10.1055/s-2007-971176 (2007).

25 Saudou, F. & Humbert, S. The Biology of Huntingtin. *Neuron* **89**, 910-926, doi:10.1016/j.neuron.2016.02.003 (2016).

26 Duyao, M. P. *et al.* Inactivation of the mouse Huntington's disease gene homolog Hdh. *Science* **269**, 407-410, doi:10.1126/science.7618107 (1995).

27 Hodges, A. *et al.* Regional and cellular gene expression changes in human Huntington's disease brain. *Hum Mol Genet* **15**, 965-977, doi:10.1093/hmg/ddl013 (2006).

28 Bezprozvanny, I. & Hayden, M. R. Deranged neuronal calcium signaling and Huntington disease. *Biochem Biophys Res Commun* **322**, 1310-1317, doi:10.1016/j.bbrc.2004.08.035 (2004).

29 Cummings, D. M. *et al.* Aberrant cortical synaptic plasticity and dopaminergic dysfunction in a mouse model of Huntington's disease. *Hum Mol Genet* **15**, 2856-2868, doi:10.1093/hmg/ddl224 (2006).

30 Bedard, C. *et al.* Serotonin and dopamine striatal innervation in Parkinson's disease and Huntington's chorea. *Parkinsonism Relat Disord* **17**, 593-598, doi:10.1016/j.parkreldis.2011.05.012 (2011).

31 Stoy, N. *et al.* Tryptophan metabolism and oxidative stress in patients with Huntington's disease. *J Neurochem* **93**, 611-623, doi:10.1111/j.1471-4159.2005.03070.x (2005).

32 Karasinska, J. M. & Hayden, M. R. Cholesterol metabolism in Huntington disease. *Nat Rev Neurol* **7**, 561-572, doi:10.1038/nrneurol.2011.132 (2011).

33 Chalmers, R. J., Johnson, R. H., Keogh, H. J. & Nanda, R. N. Growth hormone and prolactin response to bromocriptine in patients with Huntington's chorea. *J Neurol Neurosurg Psychiatry* **41**, 135-139, doi:10.1136/jnnp.41.2.135 (1978).

34 Almeida-Santos, A. F., Kangussu, L. M. & Campagnole-Santos, M. J. The Renin-Angiotensin System and the Neurodegenerative Diseases: A Brief Review. *Protein Pept Lett* **24**, 841-853, doi:10.2174/0929866524666170822120258 (2017).

35 Kumar, L. & M, E. F. Mfuzz: a software package for soft clustering of microarray data. *Bioinformation* **2**, 5-7, doi:10.6026/97320630002005 (2007).

36 Zhang, Z. *et al.* Zfhx3 is required for the differentiation of late born D1-type medium spiny neurons. *Exp Neurol*, 113055, doi:10.1016/j.expneurol.2019.113055 (2019).
37 Keilani, S. *et al.* Egr-1 induces DARPP-32 expression in striatal medium spiny neurons via a conserved intragenic element. *J Neurosci* **32**, 6808-6818, doi:10.1523/JNEUROSCI.5448-11.2012 (2012).
38 Gerfen, C. R. *et al.* D1 and D2 dopamine receptor-regulated gene expression of striatonigral and striatopallidal neurons. *Science* **250**, 1429-1432, doi:10.1126/science.2147780 (1990).
39 Ouma, W. Z., Pogacar, K. & Grotewold, E. Topological and statistical analyses of gene regulatory networks reveal unifying yet quantitatively different emergent properties. *PLoS Comput Biol* **14**, e1006098, doi:10.1371/journal.pcbi.1006098 (2018).
40 Sjostedt, E. *et al.* An atlas of the protein-coding genes in the human, pig, and mouse brain. *Science* **367**, doi:10.1126/science.aay5947 (2020).
41 Pinero, J. *et al.* The DisGeNET knowledge platform for disease genomics: 2019 update. *Nucleic Acids Res* **48**, D845-D855, doi:10.1093/nar/gkz1021 (2020).
42 Nepusz, T., Yu, H. & Paccanaro, A. Detecting overlapping protein complexes in protein-protein interaction networks. *Nat Methods* **9**, 471-472, doi:10.1038/nmeth.1938 (2012).
43 Intson, K. *et al.* Progressive neuroanatomical changes caused by Grin1 loss-of-function mutation. *Neurobiol Dis* **132**, 104527, doi:10.1016/j.nbd.2019.104527 (2019).
44 Goodspeed, A., Jean, A. & Costello, J. C. A Whole-genome CRISPR Screen Identifies a Role of MSH2 in Cisplatin-mediated Cell Death in Muscle-invasive Bladder Cancer. *Eur Urol* **75**, 242-250, doi:10.1016/j.eururo.2018.10.040 (2019).
45 Hassen, S. *et al.* Interdependence of DNA mismatch repair proteins MLH1 and MSH2 in apoptosis in human colorectal carcinoma cell lines. *Mol Cell Biochem* **412**, 297-305, doi:10.1007/s11010-015-2636-3 (2016).
46 Oben, K. Z. *et al.* Radiation Induced Apoptosis of Murine Bone Marrow Cells Is Independent of Early Growth Response 1 (EGR1). *PLoS One* **12**, e0169767, doi:10.1371/journal.pone.0169767 (2017).
47 Stenman, J., Toresson, H. & Campbell, K. Identification of two distinct progenitor populations in the lateral ganglionic eminence: implications for striatal and olfactory bulb neurogenesis. *J Neurosci* **23**, 167-174 (2003).
48 Zhao, Y. *et al.* The LIM-homeobox gene Lhx8 is required for the development of many cholinergic neurons in the mouse forebrain. *Proc Natl Acad Sci U S A* **100**, 9005-9010, doi:10.1073/pnas.1537759100 (2003).
49 Fragkouli, A., van Wijk, N. V., Lopes, R., Kessaris, N. & Pachnis, V. LIM homeodomain transcription factor-dependent specification of bipotential MGE progenitors into cholinergic and GABAergic striatal interneurons. *Development* **136**, 3841-3851, doi:10.1242/dev.038083 (2009).
50 Pevny, L. H., Sockanathan, S., Placzek, M. & Lovell-Badge, R. A role for SOX1 in neural determination. *Development* **125**, 1967-1978 (1998).
51 Baba, Y., Watabe, Y., Sagara, H. & Watanabe, S. Sall1 plays pivotal roles for lens fiber cell differentiation in mouse. *Biochem Biophys Res Commun* **512**, 927-933, doi:10.1016/j.bbrc.2019.03.098 (2019).
52 Kuroda, T. *et al.* SALL3 expression balance underlies lineage biases in human induced

pluripotent stem cell differentiation. *Nat Commun* **10**, 2175, doi:10.1038/s41467-019-09511-4 (2019).

53 Vogt, D. *et al.* Lhx6 directly regulates Arx and CXCR7 to determine cortical interneuron fate and laminar position. *Neuron* **82**, 350-364, doi:10.1016/j.neuron.2014.02.030 (2014).

54 Maere, S., Heymans, K. & Kuiper, M. BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinformatics* **21**, 3448-3449, doi:10.1093/bioinformatics/bti551 (2005).

55 Margolin, A. A. *et al.* ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. *BMC Bioinformatics* **7 Suppl 1**, S7, doi:10.1186/1471-2105-7-S1-S7 (2006).

56 Han, J. D. *et al.* Evidence for dynamically organized modularity in the yeast protein-protein interaction network. *Nature* **430**, 88-93, doi:10.1038/nature02555 (2004).

57 Rivera-Oliver, M. & Diaz-Rios, M. Using caffeine and other adenosine receptor antagonists and agonists as therapeutic tools against neurodegenerative diseases: a review. *Life Sci* **101**, 1-9, doi:10.1016/j.lfs.2014.01.083 (2014).

58 Hu, L. Z. *et al.* EPIC: software toolkit for elution profile-based inference of protein complexes. *Nat Methods* **16**, 737-742, doi:10.1038/s41592-019-0461-4 (2019).

59 Han, H. *et al.* TRRUST v2: an expanded reference database of human and mouse transcriptional regulatory interactions. *Nucleic Acids Res* **46**, D380-D386, doi:10.1093/nar/gkx1013 (2018).

60 Liu, W., Lagutin, O. V., Mende, M., Streit, A. & Oliver, G. Six3 activation of Pax6 expression is essential for mammalian lens induction and specification. *EMBO J* **25**, 5383-5395, doi:10.1038/sj.emboj.7601398 (2006).

61 Liu, Z. *et al.* Sp9 Regulates Medial Ganglionic Eminence-Derived Cortical Interneuron Development. *Cerebral Cortex*, doi:10.1093/cercor/bhy133 (2018).

62 Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139-140, doi:10.1093/bioinformatics/btp616 (2010).

63 Kohl, M., Wiese, S. & Warscheid, B. Cytoscape: software for visualization and analysis of biological networks. *Methods Mol Biol* **696**, 291-303, doi:10.1007/978-1-60761-987-1_18 (2011).

64 Huynh-Thu, V. A., Irrthum, A., Wehenkel, L. & Geurts, P. Inferring regulatory networks from expression data using tree-based methods. *PLoS One* **5**, doi:10.1371/journal.pone.0012776 (2010).

Figure legends

Fig.1 Overview of the experimental design.

First, time-points of striatum mRNA measured during mouse development using RNA-Seq. Then, the raw transcriptome data were pre-processed to obtain the DEGs. The DEGs were divided into two parts: PEGs (the genes expressed in only one stage) and CEGs (the genes expressed in at least two stages). Next, the functional TF networks of CEGs were predicted based on co-expression analysis and machine learning. Additionally, the homologous and disease genes were annotated from existing databases. In addition, the functional significance of these TF networks was tested. Subsequently, the hub TFs in the related TF networks were predicted based on the ARACNE algorithm. Finally, CKO mice were constructed to verify the hub TFs and infer their GRN.

Fig. 2 Striatal development and correlation analyses with HD

a Expression patterns of DEGs at different stages. The number of genes and TFs in each stage are shown on the right.

b Pattern of striatal development in mice.

c Histogram of DEGs (FDR < 0.05) in the striatum from P14 to P60. The top 10 significant KEGG pathways of these DEGs are displayed.

d Two-tailed hypergeometric tests were conducted to determine whether there is a significant overlap of altered genes on a specific pathway between striatal development and HD. Correlations on seven pathways between striatal development and HD.

e The number of significantly altered and overlapping genes on seven signaling pathways in striatal development and HD. The genes (FDR < 0.05) differentially expressed in striatal development and HD (adjusted P < 0.05)

f Temporal expression patterns of 57 overlapping genes across 5 time-points.

Fig. 3 Co-expression analysis of differentially expressed continuous genes

a Hierarchical clustering dendrogram based on the dynamic hybrid tree cut algorithm shows co-expression modules of continuous genes that are colour coded. The co-expressed genes are rearranged, and fuzzy c-means clustering identifies 2 distinct temporal patterns of each co-expression module. The x axis represents five developmental points, while the y axis represents log2-transformed, normalized intensity ratios in each stage. The number of genes and TFs in each module are shown on the right.

b Expression pattern of reported striatal development genes at 5 time-points.

c The expression patterns of striatal development genes were verified by RNA in situ hybridization at 5 time-points. Scale bars: 100 μm in c.

Fig. 4 Predicted potential functional TF networks based on machine learning

a Expression patterns of 740 differentially expressed continuous TFs from different co-expression modules.

b Receiver operating characteristic of differentially expressed continuous TFs from each co-expression modules. The AUC is used to assess the quality of machine learning. The AUC of the diagonal is 0.5.

c Functional significance distribution of predicted TF networks and random TF networks.

d Schematic of the global atlas of inferred functional TF networks in striatal development.

e Example TF networks with different functions of known and unknown subunits. The width of the edges increases with increasing correlation score. The shape of the node represents whether it is a homologous TF (circular node for homologous to humans, triangle node for unique in mouse). The colour of the node border represents whether the TFs are associated with nervous system–associated disease (black border for disease TFs, grey border for non-disease TFs). The black node indicates potential functional TFs. Expression patterns of representative TF networks are shown on the right by heatmap.

Fig. 5 Novel functions of Six3 predicted from TF networks

a Part of the BiNGO result for a TF network, as visualized in Cytoscape. Red categories are the most significantly overrepresented. Yellow nodes are not significantly overrepresented, which include orange nodes in the context of the GO hierarchy. The area of a node is proportional to the number of genes in the test network annotated to the corresponding GO category.

b Topological representation of the reconstructed “regulation of neuron apoptotic process” TF regulatory network using ARACNE. TFs are represented as nodes and inferred interactions as edges. Grey nodes indicate the known functional TFs. Potential functional TFs are coloured orange. The degree of orange nodes is shown on the right.

c Significant increase in expression of cleaved caspase-3 in Six3-CKO striatum compared to controls at P0. Scale bars: 100 μm in c.

d Histogram showing that the number of cleaved caspase-3^{+} cells was increased in Six3-CKO mice. (Student's t-test, ***$P < 0.001$, n=3 mice per group, mean ± SEM). Lv: lateral ventricle. Str: striatum.

e Histogram of DEGs in the LGE at P0 between control and Six3-CKO mice. The 16 DEGs associated with “regulation of neuron apoptotic process” are shown.

f A GRN of these “regulation of neuron apoptotic process” genes was inferred by Genie3. The edge direction represents the regulatory interaction from regulatory genes to target genes. Node size and colour are proportional to the number of genes (in-degree and out-degree).

g Key downstream genes that are candidates for regulating the neuronal apoptotic process. The downstream genes shown are all significantly changed in Six3-CKO mice versus wild-type mice. Genes are coloured by their importance (degree) in the “regulation of neuron apoptotic process” GRN, and the font size increases with the degree of node.

Fig. 6 Meis2 regulates neuronal apoptotic processes in the striatum

a The generation of the Meis2^{Flox} allele. The Meis2 gene-targeting vector was designed to delete exon 8 of Meis2.

b Deletion of the Meis2 gene was confirmed at the mRNA level by in situ hybridization.

c Significant increase in expression of cleaved caspase-3 in Meis2-CKO striatum compared to controls at P0. Scale bars: 100 μm in b and c.

d Histogram showing that cleaved caspase-3^{+} cells were increased in Meis2-CKO at P0. (Student's t-test, ***$P < 0.001$, n = 3 mice per group, mean ± SEM). Cx: cerebral cortex. Str: striatum.

e Histogram of DEGs in the LGE between control and Meis2-CKO mice at P0. The 18 DEGs associated with “regulation of neuron apoptotic process” are shown.

f A GRN of these “regulation of neuron apoptotic process” genes was inferred by Genie3. The edge direction represents the regulatory interaction from regulatory genes to target genes. Node size and colour are proportional to the number of genes (in-degree and out-degree).

g Key downstream genes that are candidates for regulating the neuronal apoptotic process. The downstream genes shown are all significantly changed in Meis2-CKO mice versus control mice. Genes are coloured by their importance (degree) in the “regulation of neuron apoptotic process” GRN, and the font size increases with the degree of node.

Figures

Figure 1

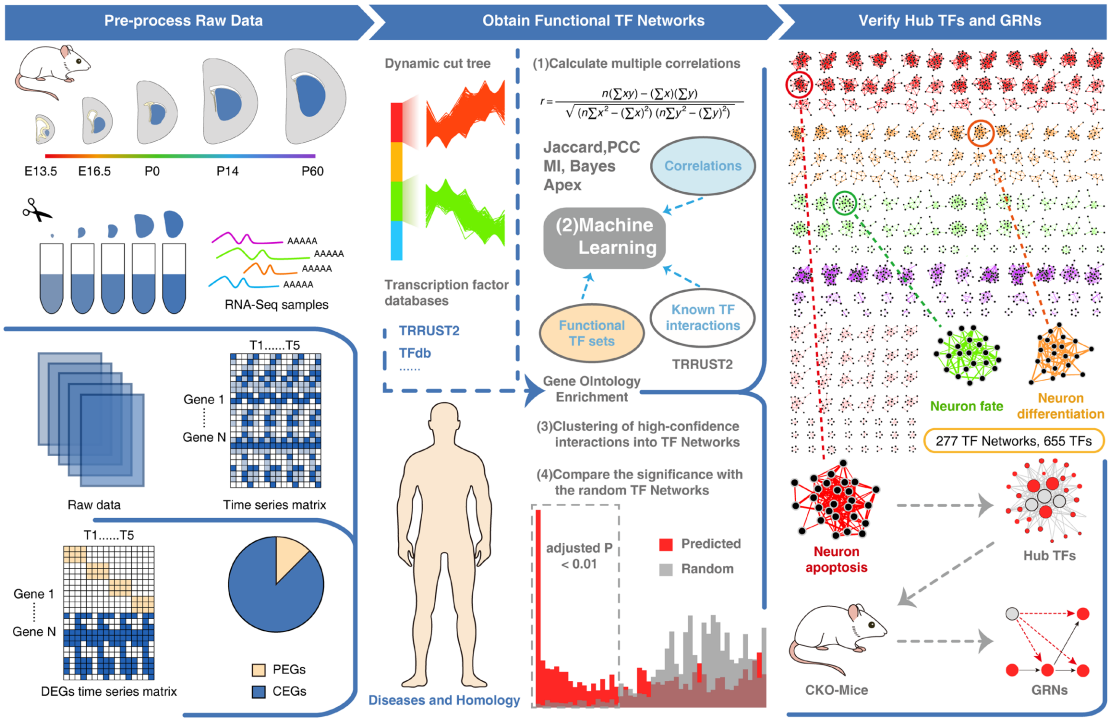
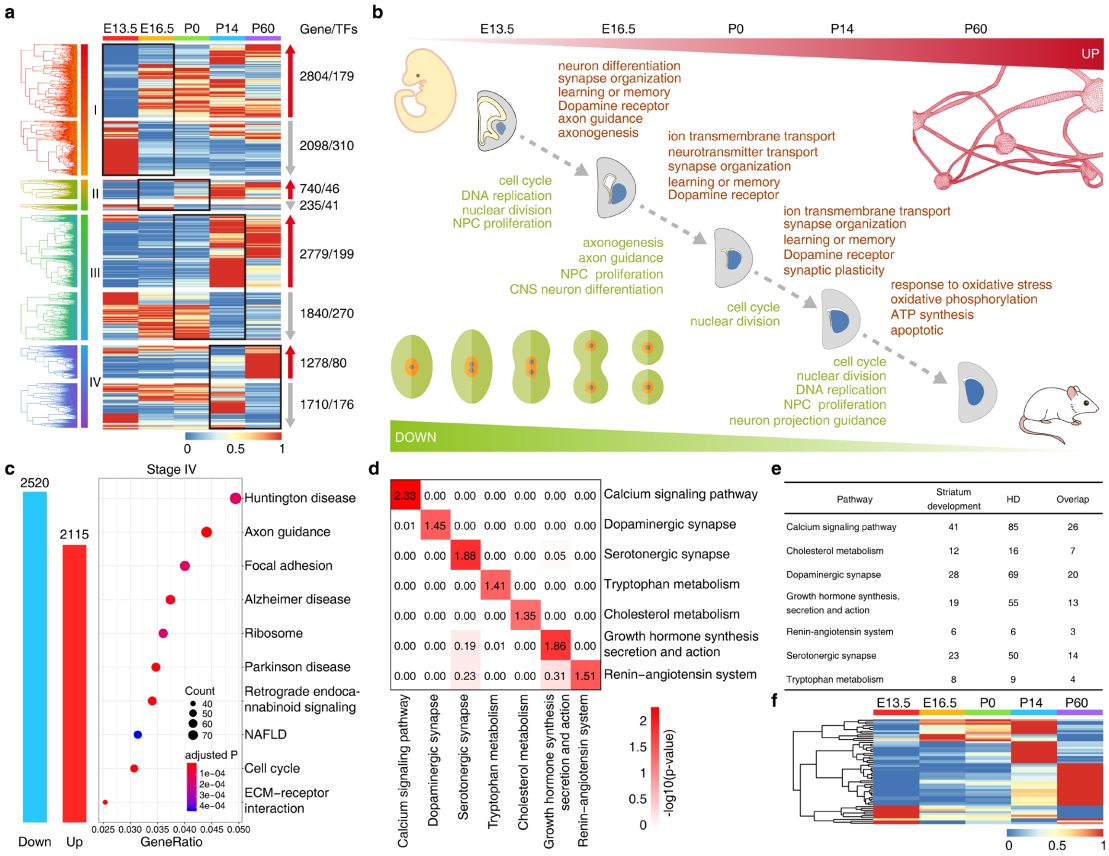


Figure 2



Pathway	Striatum development	HD	Overlap
Calcium signaling pathway	41	85	26
Cholesterol metabolism	12	16	7
Dopaminergic synapse	28	69	20
Growth hormone synthesis, secretion and action	19	55	13
Renin-angiotensin system	6	6	3
Serotonergic synapse	23	50	14
Tryptophan metabolism	8	9	4

Figure 3

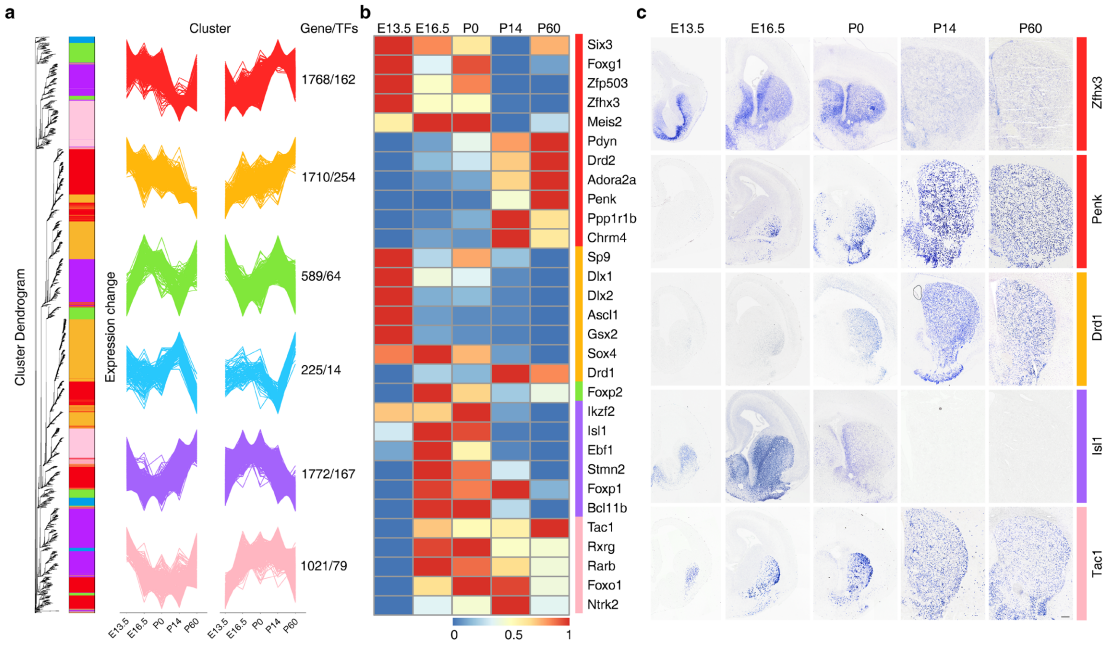

a
Cluster
Gene/TFs
Cluster Dendrogram
Expression change
1768/162
1710/254
589/64
225/14
1772/167
1021/79
E13.5
E16.5
P0
P14
P60
b
E13.5 E16.5 P0 P14 P60
Six3
Foxg1
Zfp503
Zfhx3
Meis2
Pdyn
Drd2
Adora2a
Penk
Ppp1r1b
Chrm4
Sp9
Dlx1
Dlx2
Ascl1
Gsx2
Sox4
Drd1
Foxp2
Ikzf2
Isl1
Ebf1
Stmn2
Foxp1
Bcl11b
Tac1
Rxrg
Rarb
Foxo1
Ntrk2
0
0.5
1
c
E13.5
E16.5
P0
P14
P60
Zfhx3
Penk
Drd1
Isl1
Tac1

Figure 4

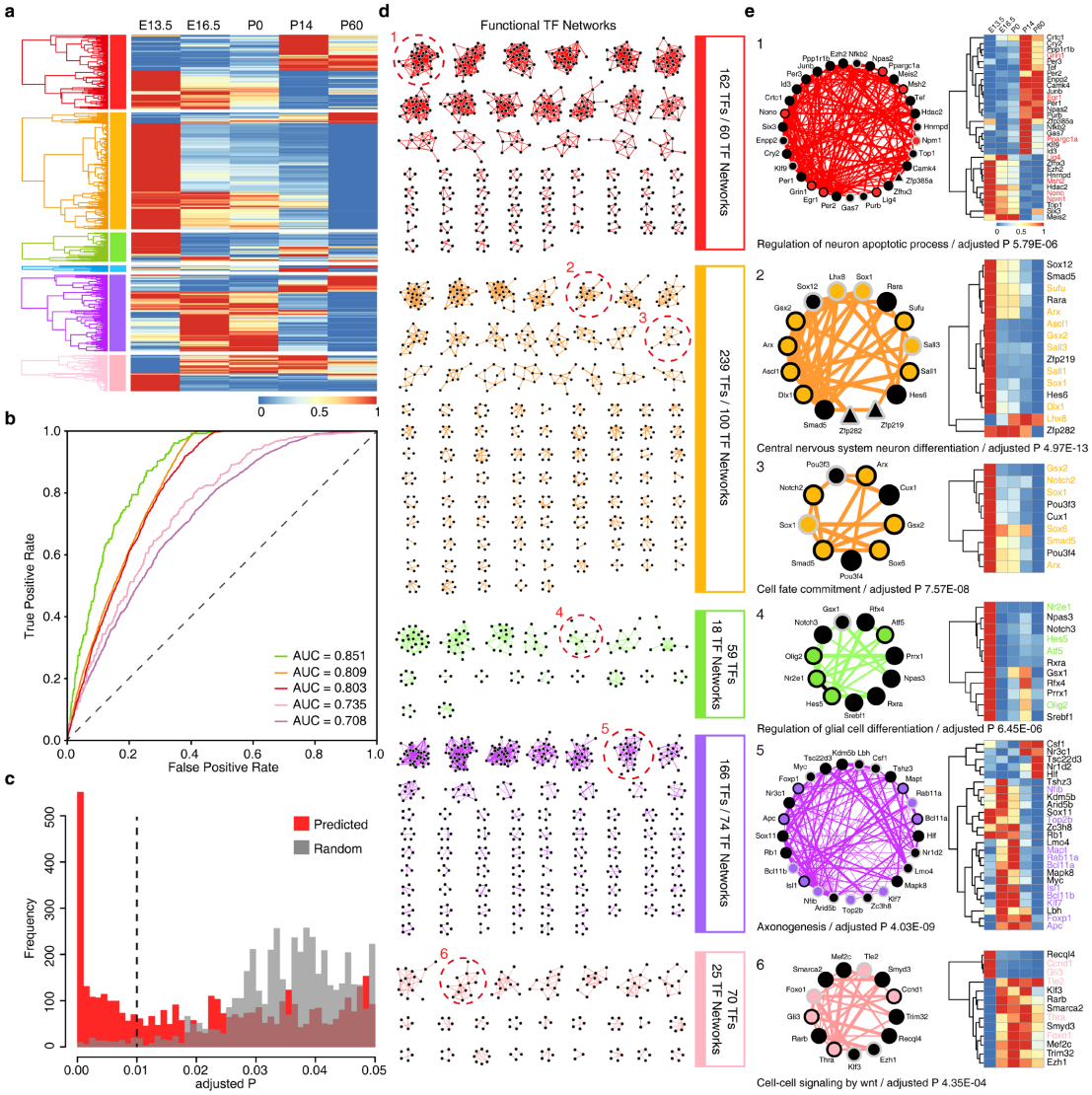

a
E13.5
E16.5
P0
P14
P60
0
0.5
1
b
True Positive Rate
False Positive Rate
AUC = 0.851
AUC = 0.809
AUC = 0.803
AUC = 0.735
AUC = 0.708
c
Frequency
adjusted P
Predicted
Random
d
Functional TF Networks
162 TFs / 60 TF Networks
239 TFs / 100 TF Networks
59 TFs
18 TF Networks
166 TFs / 74 TF Networks
70 TFs
25 TF Networks
e
Regulation of neuron apoptotic process / adjusted P 5.79E-06
Central nervous system neuron differentiation / adjusted P 4.97E-13
Cell fate commitment / adjusted P 7.57E-08
Regulation of glial cell differentiation / adjusted P 6.45E-06
Axonogenesis / adjusted P 4.03E-09
Cell-cell signaling by wnt / adjusted P 4.35E-04

Figure 5

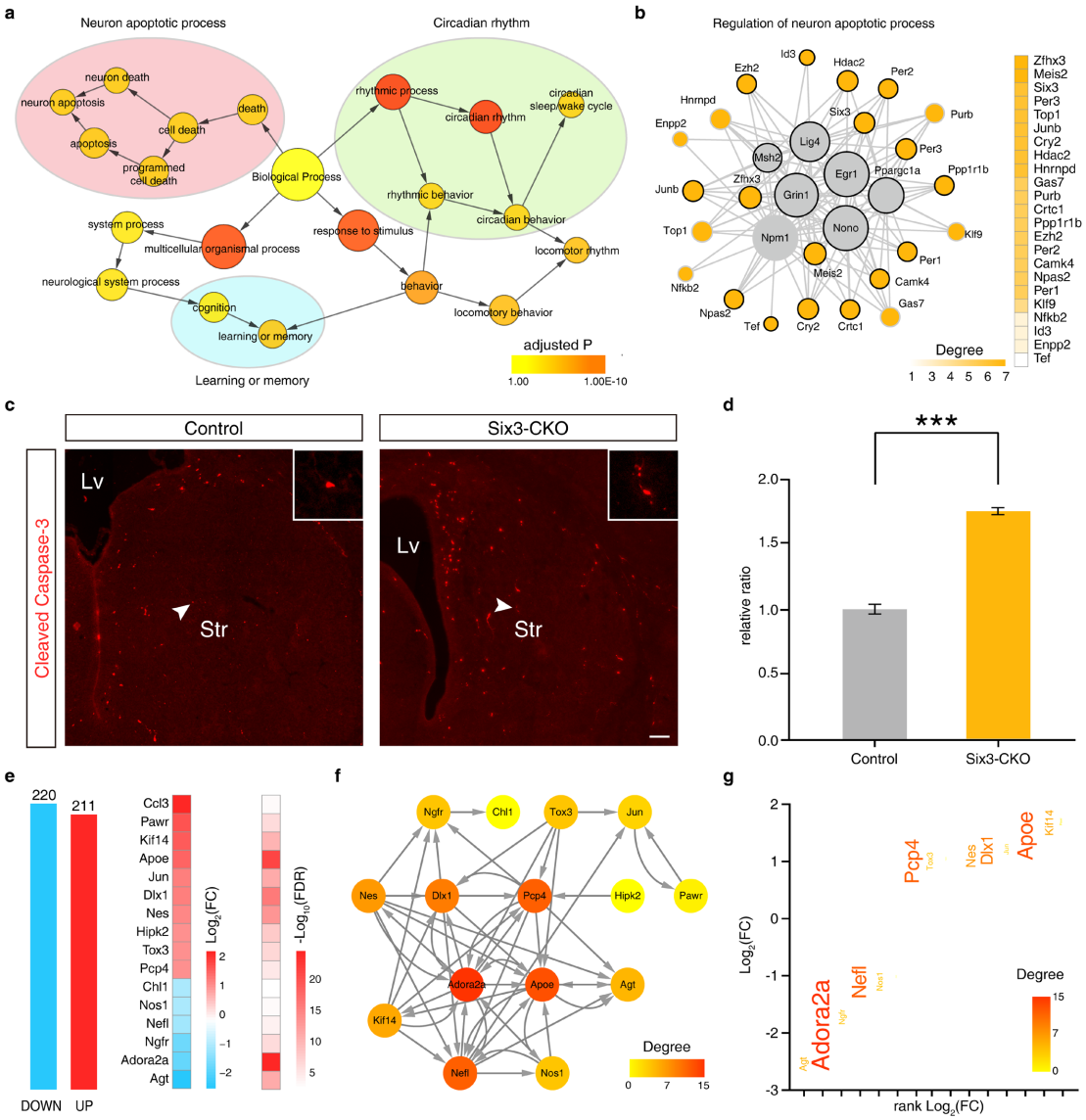
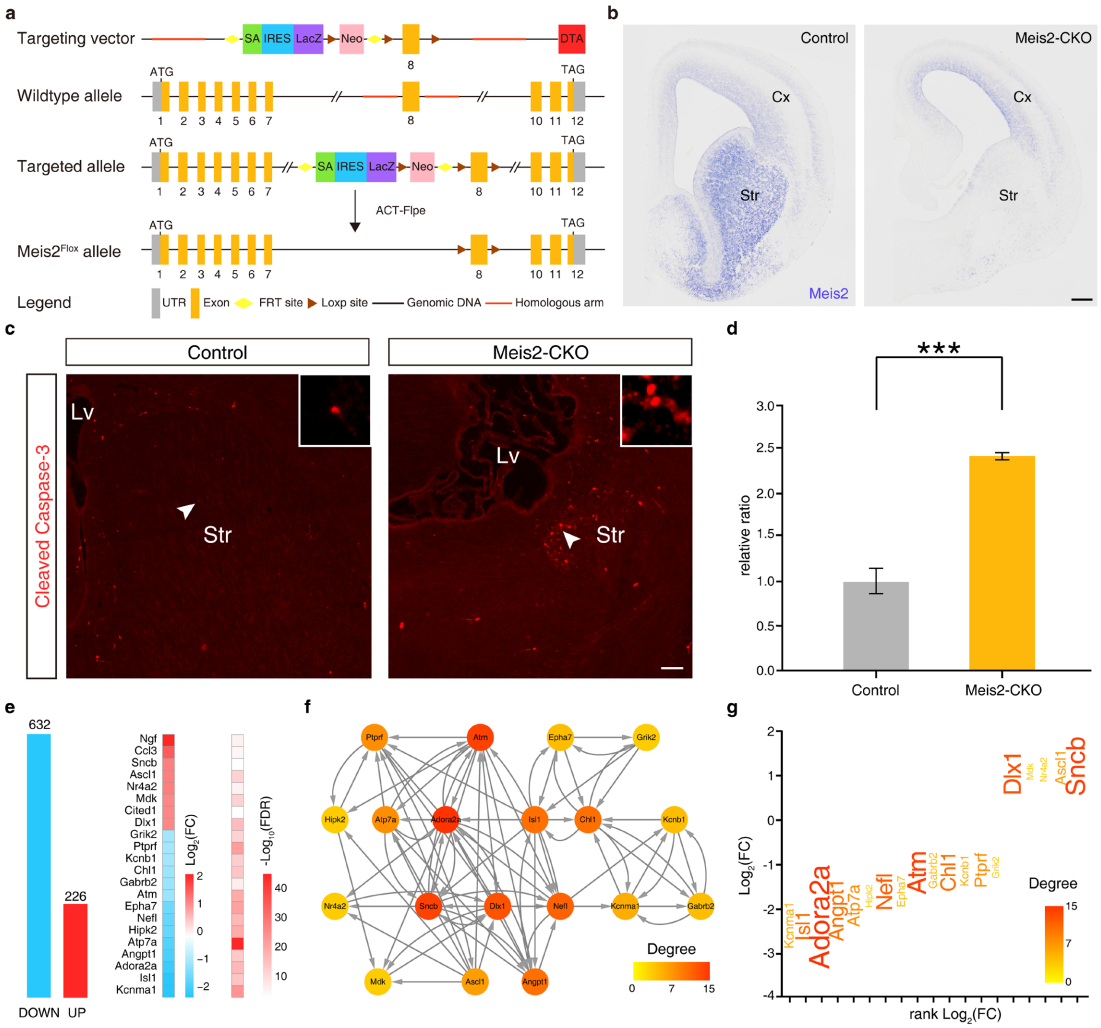

a
Neuron apoptotic process
Circadian rhythm
neuron death
neuron apoptosis
death
cell death
apoptosis
programmed cell death
Biological Process
rhythmic process
circadian rhythm
circadian sleep/wake cycle
rhythmic behavior
circadian behavior
system process
multicellular organismal process
response to stimulus
locomotor rhythm
neurological system process
behavior
cognition
locomotory behavior
learning or memory
Learning or memory
adjusted P
1.00
1.00E-10
b
Regulation of neuron apoptotic process
Id3
Ezh2
Hdac2
Per2
Hnrnpd
Six3
Purb
Enpp2
Lig4
Per3
Msh2
Egr1
Ppargc1a
Ppp1r1b
Junb
Zfhx3
Grin1
Top1
Nono
Klf9
Npm1
Per1
Meis2
Nfkb2
Camk4
Gas7
Npas2
Tef
Cry2
Crtc1
Zfhx3
Meis2
Six3
Per3
Top1
Junb
Cry2
Hdac2
Hnrnpd
Gas7
Purb
Crtc1
Ppp1r1b
Ezh2
Per2
Camk4
Npas2
Per1
Klf9
Nfkb2
Id3
Enpp2
Tef
Degree
1 3 4 5 6 7
c
Control
Six3-CKO
Cleaved Caspase-3
Lv
Str
Lv
Str
d
***
relative ratio
2.0
1.5
1.0
0.5
0.0
Control
Six3-CKO
e
220
211
DOWN
UP
Ccl3
Pawr
Kif14
Apoe
Jun
Dlx1
Nes
Hipk2
Tox3
Pcp4
Chl1
Nos1
Nefl
Ngfr
Adora2a
Agt
Log2(FC)
2
1
0
−1
−2
-Log10(FDR)
20
15
10
5
f
Ngfr
Chl1
Tox3
Jun
Nes
Dlx1
Pcp4
Hipk2
Pawr
Adora2a
Apoe
Agt
Kif14
Nefl
Nos1
Degree
0
7
15
g
Log2(FC)
2
1
0
-1
-2
-3
Pcp4
Nes
Dlx1
Apoe
Kif14
Nefl
Adora2a
Degree
15
7
0
rank Log2(FC)

Figure 6


a
Targeting vector
SA IRES LacZ
Neo
DTA
Wildtype allele
ATG
TAG
Targeted allele
ACT-Flpe
Meis2Flox allele
Legend
UTR
Exon
FRT site
Loxp site
Genomic DNA
Homologous arm
b
Control
Meis2-CKO
Cx
Str
Meis2
c
Control
Meis2-CKO
Cleaved Caspase-3
Lv
Str
d
***
relative ratio
Control
Meis2-CKO
e
632
226
DOWN
UP
Ngf
Ccl3
Sncb
Ascl1
Nr4a2
Mdk
Cited1
Dlx1
Grik2
Ptprf
Kcnb1
Chl1
Gabrb2
Atm
Epha7
Nefl
Hipk2
Atp7a
Angpt1
Adora2a
Isl1
Kcnma1
Log2(FC)
-Log10(FDR)
f
Degree
g
Log2(FC)
rank Log2(FC)
Degree