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

Lin Yang[1,2], Zihao Su[1,2], Ziwu Wang[1], Zhenmeiyu Li[1], Zicong Shang[1], Heng Du[1], Guoping Liu[1], Dashi Qi[1], Zhengang Yang[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, Shanghai Key Lab of Birth Defect, Children's Hospital, 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, Fudan University, 138 Yi Xue Yuan Road, Shanghai 200032, China.

[2]These authors contributed equally to this work.

Abstract

The striatum is structurally highly diverse, and its organ functionality critically depends on normal embryonic development. Although several studies have been conducted 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 transcriptome profile that allowed us to explore the trajectory of striatal development and identify the correlation between the striatal development and Huntington's disease (HD). Furthermore, we applied an integrative transcriptomic profiling approach based on machine learning to systematically map a global landscape of 277 transcription factor (TF) networks. Most of these TF networks are linked to biological processes, and some unannotated genes provide information about the corresponding mechanisms. For example, we found that the Meis2 and Six3 were crucial for the survival of striatal neurons, which was verified using conditional knockout (CKO) mice. Finally, we used RNA-Seq to speculate their downstream targets.

Introduction

The striatum is the major information processing centre 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]. The dysfunction of neurotransmission or neurodegeneration of the striatum results in a number of neurological disorders, including Huntington's disease (HD), Parkinson's disease (PD) and obsessive–compulsive disorders [2-4]. Thus, exploring the striatal development may provide new insights for the treatment of these diseases.

Transcription factors (TFs) and genetic regulatory networks (GRNs) are essential for the differentiation, migration and survival of striatal neurons. Currently, most studies have focused on a few genes through specific individual pathway to regulate striatal development. Based on these efforts, the key roles of several transcriptional factors (Gsx2, Dlx1/2, Sp8/9, Isl1, etc.) in controlling striatal development have been explored [5-7]. 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 through highly coordinated expression changes [8]. Thus, mapping the global landscape of TF networks during striatal development may provide some insights at the system level. Genes that belong to the similarity functional network may be co-expressed in the temporal logic and, thus, should have similar expression profiles to a certain extent [9-11]. Correlation analysis between these expression profiles have become the premise of the guilt-by-association approach, which is used to identify genes co-expressed with known genes (probes) to achieve a specific function or expression pattern [12]. This strategy of using a priori knowledge and data integration can further enhance the elucidation of gene regulatory functions [12]. 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 [13].

In the present work, we analysed 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. Furthermore, we provided the global landscape of 277 TF networks in striatal development based on machine learning, 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 neuronal apoptosis processes. Using conditional knockout (CKO) mice, we verified that Six3 and Meis2 are indeed involved in the neuron apoptotic processes 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 analysing the RNA-Seq data of the Six3 and Meis2 CKO mice, respectively. In conclusion, this study provides global TF networks, which offer valuable assets and provide direct insight into the molecular details of the striatal developmental process, enabling us to better understand the striatal development and its related diseases.

Results

Systematic Analysis of Striatal Development Based on Transcriptomes

To investigate the changes of gene expression during striatal development, we performed RNA-Seq from the early stage to the adult stage (Fig. 1). After calculating the normalized gene expression levels as the fragments per kilobase of the transcript per million mapped reads (FPKM; See Methods), we first performed a correlation analysis of the experimental data between 5 time points. We found that all data showed good consistency and a high degree of correlation (R = 0.95–0.99) between the temporally adjacent experiments (Fig. S1A). After pre-processing the raw data (See Methods), a list of 15443 genes was compiled from all 5 time points (Fig. S1B; Table S1).

To further investigate the developmental status of the striatum from E13.5 to P60, we obtained the differentially expressed genes (DEGs) (See Methods). We identified all 8861 DEGs, including 882 (10%) TFs (Fig. 2A; Table S2), and then performed biological process and KEGG pathway enrichment analysis to assign the genes to functional categories (Table S2). The functions of these DEGs were significantly enriched in biological processes, including the “regulation of cell cycle” (E13.5-E16.5), “regulation of synapse organization and axonogenesis” (E16.5-P0), “regulation of ion transmembrane transport” (P0-P14), “positive regulation of intrinsic apoptotic signalling pathway and ATP synthesis” (P14-P60) and other biological processes (Fig. S2A). The functions of these DEGs were significantly enriched in pathways including the “cell cycle and Hippo signalling pathway” (E13.5-E16.5), “axon guidance and cAMP signalling pathway” (E16.5-P0), “calcium signalling pathway and MAPK signalling pathway” (P0-P14), “Huntington disease and Alzheimer disease” (P14-P60) and other pathways (Fig. S2B). Throughout the whole process of striatal development, our results indicated that cell division gradually decreased. In contrast, neuronal differentiation, synapse organization and the dopamine signalling pathway were gradually enhanced from E13.5 to P14 (Fig. 2B). Interestingly, we found that the significantly altered genes from P14 to P60 are strongly associated with Huntington's

disease in the striatum (Fig. 2C). Therefore, our transcriptome dataset and a collected transcriptome dataset of HD (See Methods) obtained from 38 patients[14] enabled us to analyse the correlation between striatal development and HD at the mRNA level. To correlate striatal development with HD at the level 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; Table S3). Two-tailed hypergeometric statistical tests (See Methods) were carried out to determine whether there was significant overlapping of altered gene products on a specific pathway between striatal development and HD (Fig. 2E and 2F). 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.

Co-expression Analysis Identified Interrelated 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 Methods; Table S4). We used all 1257 PEGs, including 115 (9%) TFs from E13.5 to P60, and then performed biological process enrichment to assign the genes to functional categories (Fig. S3A and S3B). Our results show that the PEG biological process is in line with our delineative striatal development model in Fig. 2B. We applied weighted gene co-expression network analysis (WGCNA) [15] to obtain a system-wide understanding of groups of CEGs (7095 genes, including 740 TFs) (Table S5) whose co-expression patterns are highly correlated during striatal development. The soft-threshold power of WGCNA was defined as 13 and the scale-free topology index was > 0.85 (Fig. S4A). 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; Table S5). Then, to show the trends in gene expression, the genes from each module were grouped into

143 two trend clusters using the fuzzy c-means algorithm [16] (Fig. 3A; See Methods). Clearly,
144 each module contained not only positively correlated genes but also negatively
145 correlated genes. Next, the network heatmap plot further suggested those main
146 modules were independent from each other (Fig. S4B).

147 In addition, biological process and KEGG enrichment analysis of the gene members
148 in each module revealed distinct biological functions and pathways (Fig. S5A and S5B;
149 Table S5). Intriguingly, genes in the orange module were continuously increased or
150 decreased at the 5 time points (Fig. 3A). Genes in this module are enriched in "mitotic
151 nuclear division" and "DNA replication" (Fig. S5A and S5B), which is consistent with
152 cell division and neuronal differentiation during striatal development. Many genes are
153 essential for striatal MSN proliferation and differentiation, including Gsx2, Dlx1/2 and
154 Sp9, which were defined in this module (Fig. 3B). Genes in the red module were
155 enriched in pathways related to "synapse organization", "regulation of transmembrane
156 transporter activity", and "neurotransmitter transport", and all were related to synaptic
157 transmission (Fig. S5A). Some genes were essential for striatal development, including
158 Zfhx3, Adora2a, Ppp1r1b and Drd2, which were defined in this module (Fig. 3B).
159 Finally, we selected genes randomly and verified the expression trends of these genes
160 (Zfhx3, Penk, Drd1, Tac1 and Isl1) by in situ hybridization (Fig. 3C). Our results
161 showed that the expression trends of these genes were correct according to our co-
162 expression modules. Altogether, these results not only make it possible to understand
163 the dynamic events during striatal development but may also help to identify key
164 factors.

165

166 **Global Map of Potential Functional TF Networks Based on Machine Learning**

167 From our previous analysis, we identified transcriptional regulators that exhibited
168 clear temporal dynamics. At the core of the processes that regulate striatal
169 development and function is the transcriptional circuitry. This prompted us to focus on
170 TF expression in our time series. We identified 740 differentially expressed TFs from
171 six CEG co-expression modules (Fig. 4A; Table S5), and these TFs were used to

172 predict functional TF networks based on machine learning. Using the co-expression of
173 TFs with optimized parameter settings and known functional sets (Table S6), machine
174 learning was performed by the following steps (See Methods): correlation metrics were
175 calculated (five algorithms), machine learning predicted interactions, and high-
176 confidence interactions were clustered into functional networks (Fig. S6A). The output
177 of machine learning was well organised and ready for use in subsequent analysis to
178 predict functional TF networks (Fig. 4B). The distribution of the interaction network
179 node degree distribution followed a power-law (Fig. S6B), which has certain biological
180 significance, and this phenomenon is also widely confirmed [17]. The high proportions of
181 human homologous TFs [18] and nervous system disease TFs [19] in the mice involved in
182 these networks suggested that the subsequent analysis was likely to provide a
183 reference for human striatum studies (Fig. S6C and S6D). Afterward, we parted the
184 network using the ClusterONE algorithm to partition the TF networks for downstream
185 analysis [20]. To assess the physiological significance of the putative striatum functional
186 networks, we analysed the TF networks for coherent biological functions based on GO
187 annotations. We found that almost all of the TF networks in striatal development were
188 significantly enriched for associations with essential processes (Table S7). To prove
189 that the predicted TF networks obtained through machine learning have more
190 biological significance than just co-expression analysis, we used the method of
191 comparison with a random model to assess the accuracy of the predicted TF networks
192 (See Methods). For each module, which was measured as the biological process
193 enrichment adjusted p-value distribution, there was markedly higher significance for
194 the function of putative TF networks compared to the same number of randomized
195 networks of TFs in the network that were not predicted (Fig. 4C and S7). Therefore,
196 277 putative TF networks were identified from five co-expression modules (Fig. 4D;
197 Table S7).

198 Afterward, we roughly predicted some novel TF functions. For example, we
199 predicted that TF network 1 was enriched in the “neural apoptosis process” with high
200 significance (adjusted P 9.85E-06) (Fig. 4E). Likewise, we predicted that TF network 2

201 was enriched in “central nervous system neuron differentiation” with high significance
202 (adjusted P 4.97E-13) (Fig. 4E). In this TF network, these TFs (Ascl1, Gsx2, Dlx1, Lhx8,
203 Sox1, Sufu, Arx, Sall1 and Sall3) have been reported to be associated with cell
204 differentiation [5,21-29]. We also predicted other functional TF networks enriched in “cell
205 fate commitment”, “axonogenesis”, and “cell-cell signalling by wnt” (Fig. 4E). These
206 functional TF networks can provide us with rich treasures for predicting novel functional
207 genes, which will be helpful for researchers carrying out work related to the striatum,
208 including exploring the core transcriptional regulators and constructing transcriptional
209 regulatory networks.
210
211 **CKO of putative TFs leads to abnormal apoptosis in striatal neurons**
212 Neuronal apoptosis is very important in the development of the nervous system, and
213 the striatum is no exception. Abnormal neuronal apoptosis leads to many neurological
214 diseases, such as HD and Parkinson's disease, which are closely related to the
215 striatum. Therefore, it is necessary to explore neuronal apoptosis during striatal
216 development, which will provide new insights and ideas for researchers. Next, we
217 mainly verified the predicted neuronal apoptosis-related TFs to further confirm that our
218 global landscape of TF networks was reliable. The TF network mentioned above (Fig.
219 4E), is significantly enriched for biological processes including three main categories
220 “neuron apoptotic process” (Fig. 5A) based on the GO term relationships using
221 BiNGO[30]. To test this hypothesis, we selected several core factors: Zfhx3, Six3 and
222 Meis2 (Fig. 5B), in the center of the network, all of which were poorly studied in
223 “regulation of neuron apoptotic process” in the striatum. Recently, the TF Zfhx3 was
224 reported to induce abnormal cell apoptosis in the striatum during the first postnatal
225 week [31]. However, whether Six3 and Meis2 are involved in the cell apoptosis of striatal
226 MSNs is largely unknown.
227 To investigate whether Six3 plays a functional role in striatal MSN apoptosis, we
228 conditionally deleted the Six3 gene in the LGE using Dlx5/6-Cre lines (henceforth
229 described as Six3-CKO). Cre recombinase was robustly expressed in the LGE and

230 Six3 was removed from the striatum [32]. Next, we performed immunostaining of cleaved
231 Caspase-3 at P0. Our results showed that the cleaved Caspase-3 positive cells were
232 increased approximately 2-fold in the Six3-CKO mice compared with their littermate
233 controls (Fig. 5C, D). To investigate the potential mechanism by which Six3 affects
234 neuronal apoptosis in the striatum, we performed RNA-Seq analysis at P0 between
235 Six3-CKO mice and their littermate controls. Our RNA-Seq data revealed that
236 approximately 431 DEGs in the Six3-CKO mice (220 downregulated and 211
237 upregulated RNA expression) and then performed biological process enrichment of
238 these DEGs to functional categories (Fig. 5E, Supplementary Table 8), suggesting that
239 Six3 plays a key role in regulating neuronal apoptosis of the LGE at P0 (regulation of
240 neuron apoptotic process/ adjusted P: 5.45E-05/ 16 DEGs). To further explore the
241 potential hub downstream factors for regulating neuronal apoptosis of Six3, we
242 constructed the GRN of 16 significantly changed genes using Genie3 (see
243 Supplementary Methods), which predicted the directed regulatory links from genes to
244 genes (Fig. 5F). To determine which genes might be controlled by Six3 in the striatum,
245 we also measured the static network parameter of this GRN. Adora2a was thought to
246 have high pleiotropic effects and play essential roles in regulating neuronal apoptosis
247 (Fig. 5G). Changes in Adora2a were expected to have larger biological impacts than
248 those occurring in genes in the network periphery. This suggests that Adora2a may be
249 a potential hub downstream factor for Six3 to regulate neuronal apoptosis. Adora2a is
250 highly expressed in D2-type MSNs and is involved in cell apoptosis[33,34]. These findings
251 suggest that Six3 regulated striatal MSN apoptosis by regulating the expression of
252 Adora2a.

253 To further verify the rationality of our global landscape of TF networks and confirm
254 whether Meis2 was involved in striatal MSN apoptosis, we designed the Meis2
255 targeting vector to generate the Meis2^{Flox} allele through the strategy of homologous
256 recombination (Fig. 6A). To analyze whether Meis2 plays a functional role in MSN
257 apoptosis, we used Dlx5/6-Cre to conditionally knockout Meis2 (henceforth described
258 as Meis2-CKO) in the striatum. Our results showed that Meis2 mRNA expression was

259 completely lost in the LGE or striatum in Meis2-CKO mice compared with that in their
260 controls by in situ hybridization (Fig. 6B). Note that the expression of Meis2 in the
261 striatum was completely lost but the expression of Meis2 in the cortex remained. Next,
262 we performed immunostaining of cleaved Caspase-3 at P0. There were more cleaved
263 Caspase-3 positive cells in the Meis2-CKO mice than in their controls (Fig. 6C, D). In
264 order to investigate the potential mechanism of Meis2 affecting neuronal apoptosis in
265 the LGE or striatum, we performed RNA-Seq analysis at E16.5 between Meis2-CKO
266 mice and their littermate controls. Our RNA-Seq revealed that approximately 858
267 DEGs in Meis2-CKO mice (632 downregulated and 226 upregulated RNA expression)
268 and then performed biological process enrichment of these DEGs to functional
269 categories (Fig. 6E; Supplementary Table 9), suggesting that Meis2 plays a key role
270 in neuronal apoptosis of the striatum (regulation of neuron apoptotic process/ adjusted
271 P: 9.65E-05/ 22 DEGs). To further explore the potential hub downstream factors for
272 regulating neuronal apoptosis of Meis2, we constructed the GRN of 22 significantly
273 changed genes using Genie3 (see Supplementary Methods), which predicted the
274 directed regulatory links from genes to genes (Fig. 6F). To determine which genes
275 might be controlled by Meis2 in the striatum, we also measured the static network
276 parameter of this GRN. Isl1 and Adora2a were thought to have high pleiotropic effects
277 and play essential roles in regulating neuronal apoptosis (Fig. 6G). Changes in Isl1 or
278 Adora2a were expected to have larger biological impacts than those occurring in genes
279 in the network periphery. Our results suggest that Isl1 and Adora2a may be potential
280 downstream factors for Meis2 to regulate the neuronal apoptosis. In fact, Isl1 and
281 Adora2a have been reported to be expressed in D1-type and D2-type MSNs and
282 regulate cell apoptosis in the striatum[6,33,34]. Altogether, these results indicated that the
283 transcriptional factors Six3 and Meis2 are the hub genes in the transcriptional network
284 for the survival of striatal MSNs.
285

286 **Discussion**

287 In this study, we presented the panoramic overview of the developing mouse
288 striatum from embryonic stage to adult. Quantitative analysis of 15443 gene products
289 allows us to explore the trajectory of striatal development and identify the correlation
290 between the striatal development and HD. More importantly, we provided the global
291 landscape of 277 TF networks based on co-expression analysis and machine learning
292 models. Furthermore, we identified the hub TFs Six3 and Meis2 which are involved in
293 regulating the apoptosis of striatal neurons. Finally, using conditional knockout (CKO)
294 mice and RNA-Seq data, we verified that Six3 and Meis2 indeed regulated neuronal
295 apoptosis and inferred their downstream targets in the striatal development stages.
296 Therefore, our study provides a resource to understand striatal development and its
297 connection to striatal disease.

298 During mouse striatal development, our preliminary bioinformatic analysis
299 suggested that cell division gradually decreased, but neuronal differentiation, synapse
300 organization and the dopamine signalling pathway gradually increased. It is worth
301 mentioning that the DEGs in the striatum from P14 to P60 are closely related to HD.
302 This phenomenon inspired us to investigate whether there is experimental evidence
303 for a connection between striatal development and HD. By comparing the mRNA
304 expression of striatal development with that of HD [14], we found that a noticeable
305 number of pathway regulators in HD are also differentially expressed during striatal
306 development. Moreover, we identified statistically significant networks of overlapping
307 genes in seven pathways between striatal development and HD, which suggests that
308 HD regains some features of striatum embryonic development but loses some basic
309 functions of the adult striatum and that HD is likely a consequence of the dysregulation
310 of master regulators and pathways.

311 The main motivation behind this study was the identification of networks of co-
312 expressed TFs, with an inherent expression distribution corresponding to discrete
313 biological processes associated with striatal development. Firstly, individual pair-wise
314 TF associations were scored based on the temporal expression profile similarity

measured during striatal development. Next, we used an integrative computational scoring procedure to derive TF interactions. The support vector machine learning classifier used was trained on correlation scores obtained for reference annotated TF sets [35], and we combined all of the input co-expression data with previously published TF interaction evidence [36]. Measurements of the overall performance showed a high area under the curve for 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, which were highly co-expressed in the LGE during the development of the striatum, along with their co-expressed targets. Deletion of the expression of these genes leads to abnormal differentiation of striatal MSNs [5,22,28]. 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. We generated Meis2-CKO mice for the first time. Using Meis2-CKO mice to perform RNA-Seq, we predicted that Isl1 and Adora2a were regulated by Meis2 in regulating the neuronal apoptosis. These findings were supported by previous reports [6,33,34]. Moreover, Using the same approach, we speculated that Six3 regulated the survival of striatal neurons by regulating the expression of Adora2a. Adora2a is highly expressed in D2-type MSNs and is involved in cell apoptosis [33,34]. This raises the exciting potential to identify more critical factors from our data and deepen our understanding of the regulation of striatal development.

339 **Methods**

340 **Animals and Tissue Collection**

341 The mouse genetic background was CD1 in this study. The Dlx5/6-Cre and Six3$^{F/F}$
342 mice were previously described [21,37]. We generated Meis2$^{F/+}$ mice in this study. The
343 details of the targeting construct are shown in Fig. 6. Briefly, we first generated Meis2
344 knockout mice via the ESC targeting system. The cassette is composed of an FRT site
345 followed by a LacZ sequence and a loxP site. This first loxP site is followed by
346 neomycin under the control of the human beta-actin promoter, SV40 polyA, a second
347 FRT site and a second loxP site. A third loxP site is inserted downstream of the
348 targeted exon 8. Exon 8 is thus flanked by the loxP site. A “conditional ready” (floxed)
349 allele can be created by flp (ACT-flp) recombinase expression in mice carrying this
350 allele. Subsequent Cre expression results in knockout mice. The Mesi2$^{F/F}$ mouse
351 genotype was identified by PCR (Polymerase Chain Reaction). The primers were as
352 follows: Meis2-loxp F1: TGTCAAACATTGCCATCCAAACAGT, and Meis2-loxp F2:
353 TCATGGAGAAGACAGCGCTGGT. Mice were group-housed with approximately
354 three mice in each cage on a 12-h light, 12-h dark cycle, with ad libitum chow and
355 water. The day of vaginal plug detection was calculated as E0.5, and the day of birth
356 was considered as P0. Both males and females were used in all experiments. All
357 animal experiments described in this study were approved in accordance with
358 institutional guidelines at Fudan University Shanghai Medical College.

359 RNA-Seq analysis was performed as previously described [38]. For the temporal
360 transcriptome analysis, the whole LGE or striatum was separated from embryonic to
361 adult mice. The LGE or striatum was collected at 5 time points: E13.5, E16.5, P0, P14
362 and P60. Each time point had three biological replicates, but we obtained only two
363 successful biological replicates at P14. The whole LGE or striatum tissues were
364 washed twice with ice-cold HBSS (Hanks’ Balanced Salt Solution) to remove blood,
365 quick-frozen in liquid nitrogen and then stored at –80 °C for total RNA extraction. The
366 striatum (including the VZ, SVZ and MZ-striatum) of the Dlx5/6-Cre; Six3$^{F/F}$ (henceforth
367 described as Six3-CKO) mice and their Dlx5/6-Cre littermates (henceforth described

368 as controls) were dissected at P0 (n=3 mice, each group). The LGE or striatum
369 (including the VZ, SVZ and MZ-striatum) of the Dlx5/6-Cre; Meis2$^{F/F}$ (henceforth
370 described as Meis2-CKO) mice and their Dlx5/6-Cre littermates (henceforth described
371 as controls) were dissected at E16.5 (n=3 mice, each group).

372 **Immunohistochemistry**

373 Mice were anaesthetized with an intraperitoneal injection of ketamine (100 mg per
374 kg) and transcardially perfused with a fixative solution containing 4%
375 paraformaldehyde (PFA). The brains were post-fixed in 4% PFA at 4 °C overnight.
376 Sections of 20 μm thickness were used for immunostaining. The sections were washed
377 with 0.05 M TBS for 10 min, incubated in Triton-X-100 (0.5% in 0.05 M TBS) for 30 min
378 at room temperature (RT), and then incubated with block solution (10% donkey serum
379 +0.5% Triton-X-100 in 0.05 M TBS, pH=7.2) for 2 h at RT. Primary antibodies against
380 cleaved Caspase-3 (1:500; Cell Signaling, 9661, 43) were diluted in 10% donkey
381 serum block solution, incubated overnight at 4 °C, and then rinsed 3 times with 0.05 M
382 TBS. Secondary antibodies (from Jackson, 1:500) matching the appropriate species
383 were incubated for 3 h at RT. The fluorescently stained sections were then washed 3
384 times with 0.05 M TBS. This was followed by 4′, 6-diamidino-2-phenylindole (DAPI)
385 (Sigma, 200 ng/ml) staining for 3 min, and the sections were then cover-slipped with
386 Gel/Mount (Biomeda, Foster City, CA). Fluorescence images were collected with an
387 Olympus-BX61VS microscope using 10× and 20× objectives.

388 **In Situ RNA Hybridization**

389 In situ RNA hybridization experiments were performed using digoxigenin labelled
390 riboprobes on 20-μm frozen sections as previously described [31,32] or made from
391 cDNAs amplified by PCR using the following primers:
392 Drd1 Fwd: GGCATTTGGAGAGATGTGGCACCAG.
393 Drd1 Rev: AGATAGCCCAATACCTGTCCACGCT.
394 Meis2 Fwd: CGATGGGTTAGACAACAGCGTAGCTTCACCTG.
395 Meis2 Rev: ATGGCTTGGCAAATATGATGCATTGGGTCCATG.
396 Isl1 Fwd: ATGGGAGACATGGGCGATCC.

397 Isl1 Rev: CATGCCTCAATAGGACTGGCTACC.

398 Penk Fwd: TCGGAAGGACAGGATGTCATCA.

399 Penk Rev: CGTCAGGAGAGATGAGGTAACAAAC.

400 Tac1 Fwd: CCCCTGAACGCACTATCTATTC.

401 Tac1 Rev: TAGAGTCAAATACCGAAGTCTCAG.

402 Zfhx3 Fwd: CGATCTGGCCCAGCTCTACCA.

403 Zfhx3 Rev: CTGTAAGCCTGCGAGGGCATAG.

404 **Quantification**

405 DAPI staining was used to define the striatal area for quantification. Brain regions were
406 identified using a mouse brain atlas and sections equivalent to the following
407 coordinates: the most-rostral section, 2.19 mm; the most-caudal section, 3.03 mm, n=3
408 mice for each genotype, 4 sections per mouse). Schema chart was shown in Figure
409 S8. For quantification of Caspase-3 positive cells in entire striatum at P0, four 20-μm
410 thick coronal sections from rostral, intermediate, and caudal levels of the striatum were
411 analysed. We counted all Caspase-3 positive cells in the striatum.

412 **RNA Sequencing and Data Processing**

413 The total RNA was extracted from the LGE or striatum tissues and treated with
414 deoxyribonuclease I (DNase I). mRNA was isolated by Oligo Magnetic Beads and cut
415 into small fragments that served as templates for cDNA synthesis. Once short cDNA
416 fragments were purified, they were extended with single-nucleotide adenines, ligated
417 with suitable adapters, and amplified by PCR before they were sequenced. High-
418 throughput RNA sequencing (RNA-Seq) experiments were carried out using the
419 Illumina comprehensive next-generation sequencing (NGS) technique. The raw data
420 were filtered and processed by FastQC software (Version 0.11.5, available online at
421 the website: http://www.bioinformatics.babraham.ac.uk/projects/fastqc). Low-quality
422 RNA-Seq reads were removed if they had a Phred quality score of less than 20 or had
423 less than 50 nucleotides. The filtered reads were mapped onto the mouse reference
424 genome (GRCm38.p2.genome, released on 12/10/2013) using HISAT2 software
425 (Version 2.1.0). Assembly and quantification of the transcripts were accomplished with

426 StringTie software (Version 1.3.1) using the mouse genome annotation file as the
427 reference (gencode.vM2. annotation, available at the GENCODE website). The
428 number of FPKM was used for the measurement of the relative abundances of the
429 transcripts.

430 The Pearson correlation coefficient was calculated between biological replicates
431 with the normalized expression levels of $\log_2$ (FPKM value +1). The missing value was
432 defaulted to 0. The comparison of the two biological replicates showed that the
433 expression values between them were highly correlated (average R = 0.98). Hence,
434 the average FPKM value of the two replicates was taken as the expression level for
435 the sample at each time point. However, prior to this point, the transcription was
436 performed at least twice in biological repetitions. Afterward, to reduce the influence of
437 transcription noise, a gene was defined as expressed if its FPKM value was ≥1 at least
438 at a time point.

439 The read counts were utilized for differential expression analyses using the edgeR
440 package in R. A key feature of the edgeR package is the use of weighted likelihood
441 methods to implement a flexible empirical Bayes approach in the absence of easily
442 tractable sampling distributions [39]. Gene expression data between adjacent time points
443 (E13.5-16.5, E16.5-P0, P0-P14 and P14-P60) were used for differential analysis. The
444 false discovery rate (FDR) was controlled by the Benjamini and Hochberg (BH)
445 algorithm. Unless otherwise specified, genes with a FDR < 0.01 and a fold-change
446 (FC) > 2 or < 0.5 were judged to be differentially expressed. Also, the DEGs were
447 divided into two parts: PEGs (the genes differentially expressed between only two
448 adjacent time points) and CEGs (the genes differentially expressed between at least
449 three adjacent time points).

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

454 **Referenced Datasets**

455 We obtained microarray gene expression profiling data from the following GEO
456 datasets (http://www.ncbi.nlm.nih.gov/geo/): GSE3790, which includes human post-
457 mortem striatal tissue from 38 HD-gene-positive cases and 32 age- and sex-matched
458 controls[14] (more information in supplementary table 11). HD cases were analysed
459 according to the presence or absence of disease symptoms and Vonsattel pathological
460 grade (scale=0–4). According to the Vonsattel scale, two neuropathologists with
461 special interest and expertise in HD pathology graded the corresponding formalin-fixed
462 and paraffin-embedded caudate section of HD cases. This grading scale is based on
463 the overall pattern of caudate nucleus neuropathology, the number and proportion of
464 neurons and astrocytes. Some cases were cross-referenced to confirm the
465 consistency of the scoring procedure. DNA was extracted from all patients and controls,
466 and the CAG repeat length allele in IT15 (HD gene) was genotyped. The Affymetrix
467 microarrays, we downloaded the gene expression matrix to identify the DEGs
468 (adjusted $P < 0.05$) by using the limma package.

469 **Functional Annotation GO Term Analysis**

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

476 **Two-Tailed Hypergeometric Statistical Test**

477 Two-tailed hypergeometric statistical tests were carried out by R (the phyper function,
478 lower.tail=F) to determine whether there was significant overlapping (p-value < 0.05)
479 of the number of DEGs on a specific pathway between striatal development (P14 to
480 P60) and HD. In the application of the phyper function, the total number of genes
481 involved in the pathway was from the KEGG database.

482 **Weighted Gene Co-expression Network Analysis**

483 Weighted gene co-expression network analysis was performed using the R package

484 WGCNA [15]. The $\log_2$ (FPKM value +1) transformed data were used to generate a
485 matrix of the Pearson correlations between all pairs of genes across the samples. The
486 value β = 13 was chosen as the saturation level for a soft threshold of the correlation
487 matrix based on the criterion of approximate scale-free topology. To minimize the
488 effects of noise and spurious connections, the correlation matrix was transformed to
489 the topological overlap matrix using the "TOMsimilarity" function [15]. The topological
490 overlap matrix was then used to group highly co-expressed genes by performing
491 average linkage hierarchical clustering. The dynamic hybrid tree-cut algorithm was
492 used to cut the hierarchal clustering tree [15], and the modules were defined as the
493 branches resulting from this tree cutting. The minimum module size was defined as 30,
494 and the threshold for the merging of similar modules was defined as 0.35.

495 Then, to show trends in gene expression, the genes from each module were grouped
496 into clusters using the Mfuzz package in R with the fuzzy c-means algorithm [16]. Before
497 clustering, standardization of the expression values of every gene was carried out, so
498 that the average expression value for each gene was 0, and the standardization
499 deviation of its expression profile was 1.

500 **Processing of Machine Learning**

501 Genes that belong to the similarity functional network may be co-expressed in the
502 temporal logic and, thus, should have similar expression profiles to a certain extent.
503 The exploration of the similarity between gene expression profiles can be transformed
504 into the exploration of the correlations between them. At the same time, the
505 correlations between known functional gene sets can be taken as an internal reference.
506 Through machine learning, it is possible to explore the similar functional networks and
507 the diversity of gene functions to a certain extent. For now, machine learning of
508 correlations between substances has been used to search for potentially relevant
509 substances, such as the components of protein complexes in their natural state [35,40,41].
510 Therefore, we hope to explore similar functional gene networks and gene functional
511 diversity on a large scale to some extent.

512 We applied several methods to measure the similarity of two gene expression

513 profiles to emphasize multiple gene expression profile features. We treated each
514 expression profile as a vector consisting of the observed FPKM for a particular gene
515 across the corresponding temporal logical, and a complete co-expression experiment
516 was stored as a matrix in which the rows and columns represent genes and times,
517 respectively. To measure the co-expression profile similarity between two genes, we
518 employed various correlation metrics that ranged from simple scores to more
519 sophisticated metrics based on information theory. The Jaccard score (Jaccard)
520 computes the ratio of how often genes are expressed at the same time point and how
521 often genes are detected at all times. Thus, the Jaccard score between two genes was
522 calculated by counting the number of time points that contained both genes and
523 dividing the result by the number of time points that included at least one of the two
524 genes. The Pearson correlation coefficient (PCC) is used to measure the similarity of
525 two gene co-expression profiles. The PCC was calculated using the vector of the raw
526 FPKM. Mutual information (MI) considers both linear and nonlinear dependencies
527 between vectors. The Bayes correlation (Bayes) was proposed to process RNA-seq
528 gene expression data based on sequence counts for various genes under different
529 conditions. Here, we applied the same method for the FPKM for various genes across
530 the temporal logic. For the Apex score (Apex), most genes tend to be expressed at a
531 specific time, and thus, the time point that contains the largest amount of a particular
532 gene is typically also the most critical time for that gene. Thus, two genes are
533 considered to be more likely to have similar functions if the time points containing the
534 largest recorded amounts of these genes across all times are the same.

535 These correlation coefficients were all between 0-1 (except the Apex, which was
536 either 0 or 1). To reduce the operation time, we reserved the pairs with the maximum
537 correlation scores (gene co-expression pairs in all four different correlation measures)
538 ≥ 0.75/0.85. Afterward, the correlation vectors were input into a supervised machine
539 learning model (random forest algorithm) that was both trained to predict new gene
540 pairs and benchmarked against reference positive (annotated) gene groupings (that is,
541 the enrichment of each module of genes in the GO database) (Table S10) and negative

542 gene groupings (that is, combinations of genes with distinct enrichments in each gene
543 module). The additional supporting evidence (for example, functional interactions
544 inferred from co-expression) was integrated from public sources TRRUST2 [36], thereby
545 producing richer and more accurate interaction networks. Finally, network partitioning
546 to define gene network membership was applied by ClusterONE [20]. Afterward, to prove
547 that the co-expressed gene networks obtained through machine learning had more
548 biological significance, the functional significance of the randomly generated gene
549 networks was compared. A randomly generated gene network is a random sampling
550 of the genes involved in machine learning, and random sampling ensures that the
551 number of gene networks and the total number of genes in each gene network are
552 consistent. We conducted a total of 100 random samplings, conducted enrichment
553 analysis of the results of each random sampling, and calculated its average frequency
554 in each significance stage to exclude the contingency of random sampling.

555 **Static Network Analysis and Gene Regulatory Network Prediction**

556 The reconstructed network was imported into the Cytoscape [42] software package for
557 subsequent analyses. The network was analysed based on the topological parameters,
558 i.e., the number of nodes, diameter, radius, centralization, density, heterogeneity,
559 number of connected components, number of the shortest paths, characteristic path
560 length, and average number of neighbours, and central parameters (i.e., the node
561 degree distribution and neighbourhood connectivity distribution) using the Network
562 Analyzer plug-in.

563 To infer the gene regulatory network, given the expression profiles and the list of
564 genes, putative regulatory links from genes to all other genes were estimated using
565 Genie3 [43] with default parameters, and only the top 25% of regulatory links with a
566 prediction weight were retained for constructing a GRN. For each node (gene) in the
567 network, the in- and out-degrees represented the number of edges (regulatory links)
568 directed to and from this node, respectively.

569 **Annotation of Genes Using Existing Datasets**

570 The transcription factor annotation integrated three databases (Table S10), which

were TRRUST2, TFdb and GO. TRRUST2 (https://www.grnpedia.org/trrust/) is a manually curated database of human and mouse transcriptional regulatory networks[36]. TFdb (http://genome.gsc.riken.jp/TFdb/) is a database containing mouse transcription factor genes and their related genes. The genes are annotated in the Gene Ontology database as GO:0003700, “DNA binding TF activity”.

The disease annotation was 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 [19]. 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 (Table S10).

The homologous annotation was from the Brain Atlas (https://www.proteinatlas.org/humanproteome/brain), which explores protein expression in the mammalian brain by the visualization and integration of data from three mammalian species (human, pig and mouse) [18].

587 **Acknowledgements**

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

591

592 **Potential conflicts of interest**

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

594

595 **Funding**

596 Research grants to Z.Yang. from National Key Research and Development Program
597 of China (2018YFA0108000), National Natural Science Foundation of China (NSFC
598 31820103006, 31630032, 31425011, and 31429002), and research grant to D. Qi
599 (NSFC 81974175).

600

601 **Data availability statement**

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

604

605 **References**

606 1 Graybiel, A. M. Building action repertoires: memory and learning functions of the basal
607 ganglia. *Curr Opin Neurobiol* **5**, 733-741, doi:10.1016/0959-4388(95)80100-6 (1995).
608 2 Wichmann, T. & DeLong, M. R. Functional and pathophysiological models of the basal
609 ganglia. *Curr Opin Neurobiol* **6**, 751-758, doi:10.1016/s0959-4388(96)80024-9 (1996).
610 3 DeLong, M. R. Primate models of movement disorders of basal ganglia origin. *Trends*
611 *Neurosci* **13**, 281-285 (1990).
612 4 Gerfen, C. R. & Surmeier, D. J. Modulation of striatal projection systems by dopamine.
613 *Annu Rev Neurosci* **34**, 441-466, doi:10.1146/annurev-neuro-061010-113641 (2011).
614 5 Waclaw, R. R., Wang, B., Pei, Z., Ehrman, L. A. & Campbell, K. Distinct temporal
615 requirements for the homeobox gene Gsx2 in specifying striatal and olfactory bulb
616 neuronal fates. *Neuron* **63**, 451-465, doi:10.1016/j.neuron.2009.07.015 (2009).
617 6 Lu, K. M., Evans, S. M., Hirano, S. & Liu, F. C. Dual role for Islet-1 in promoting striatonigral
618 and repressing striatopallidal genetic programs to specify striatonigral cell identity. *Proc*
619 *Natl Acad Sci U S A* **111**, E168-177, doi:10.1073/pnas.1319138111 (2014).
620 7 Ehrman, L. A. *et al.* The LIM homeobox gene Isl1 is required for the correct development
621 of the striatonigral pathway in the mouse. *Proc Natl Acad Sci U S A* **110**, E4026-4035,
622 doi:10.1073/pnas.1308275110 (2013).
623 8 Nord, A. S., Pattabiraman, K., Visel, A. & Rubenstein, J. L. Genomic perspectives of
624 transcriptional regulation in forebrain development. *Neuron* **85**, 27-47,
625 doi:10.1016/j.neuron.2014.11.011 (2015).
626 9 Li, X. *et al.* A time-resolved multi-omic atlas of the developing mouse stomach. *Nat*
627 *Commun* **9**, 4910, doi:10.1038/s41467-018-07463-9 (2018).
628 10 Zhang, Y. *et al.* Dynamic epigenomic landscapes during early lineage specification in
629 mouse embryos. *Nat Genet* **50**, 96-105, doi:10.1038/s41588-017-0003-x (2018).
630 11 Gao, Y. *et al.* Protein Expression Landscape of Mouse Embryos during Pre-implantation
631 Development. *Cell Rep* **21**, 3957-3969, doi:10.1016/j.celrep.2017.11.111 (2017).
632 12 Wolfe, C. J., Kohane, I. S. & Butte, A. J. Systematic survey reveals general applicability of
633 "guilt-by-association" within gene coexpression networks. *BMC Bioinformatics* **6**, 227,
634 doi:10.1186/1471-2105-6-227 (2005).
635 13 Ballouz, S., Verleyen, W. & Gillis, J. Guidance for RNA-seq co-expression network
636 construction and analysis: safety in numbers. *Bioinformatics* **31**, 2123-2130,
637 doi:10.1093/bioinformatics/btv118 (2015).
638 14 Hodges, A. *et al.* Regional and cellular gene expression changes in human Huntington's
639 disease brain. *Hum Mol Genet* **15**, 965-977, doi:10.1093/hmg/ddl013 (2006).
640 15 Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network
641 analysis. *BMC Bioinformatics* **9**, 559, doi:10.1186/1471-2105-9-559 (2008).
642 16 Kumar, L. & M, E. F. Mfuzz: a software package for soft clustering of microarray data.
643 *Bioinformation* **2**, 5-7, doi:10.6026/97320630002005 (2007).
644 17 Ouma, W. Z., Pogacar, K. & Grotewold, E. Topological and statistical analyses of gene
645 regulatory networks reveal unifying yet quantitatively different emergent properties. *PLoS*
646 *Comput Biol* **14**, e1006098, doi:10.1371/journal.pcbi.1006098 (2018).
647 18 Sjostedt, E. *et al.* An atlas of the protein-coding genes in the human, pig, and mouse brain.
648 *Science* **367**, doi:10.1126/science.aay5947 (2020).

19 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).
20 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).
21 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).
22 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).
23 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).
24 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).
25 Pevny, L. H., Sockanathan, S., Placzek, M. & Lovell-Badge, R. A role for SOX1 in neural determination. *Development* **125**, 1967-1978 (1998).
26 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).
27 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).
28 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).
29 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).
30 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).
31 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).
32 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).
33 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).
34 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).
35 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).

36 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).

37 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).

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

39 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).

40 Havugimana, P. C. *et al.* A census of human soluble protein complexes. *Cell* **150**, 1068-1081, doi:10.1016/j.cell.2012.08.011 (2012).

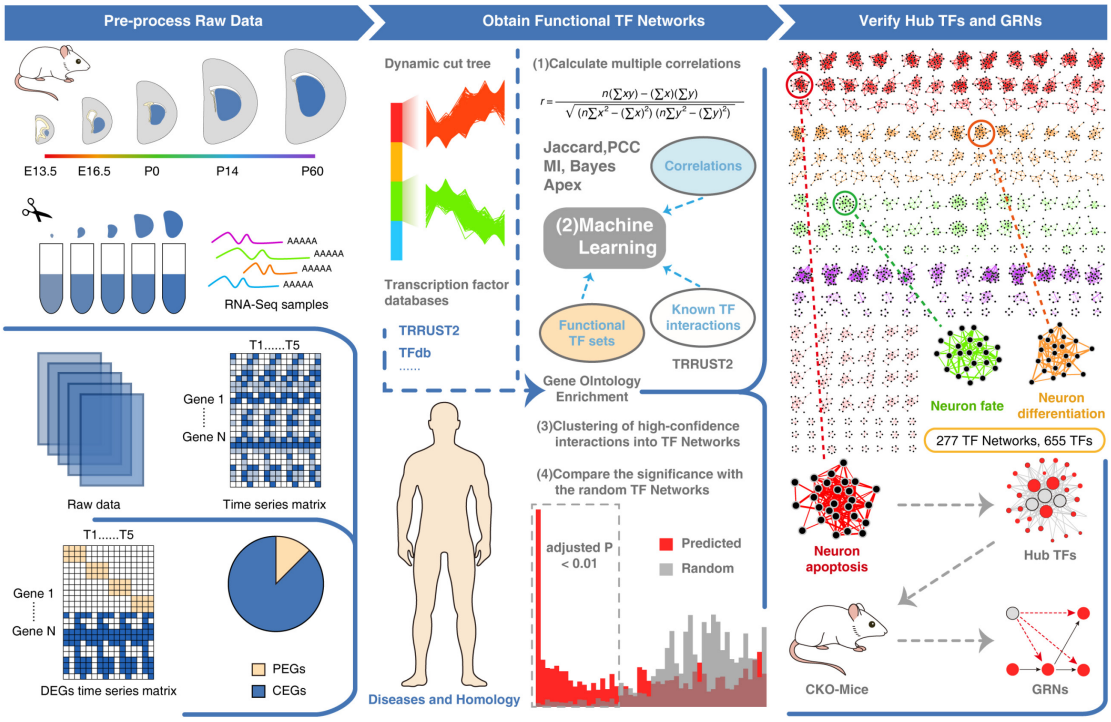
41 Wan, C. *et al.* Panorama of ancient metazoan macromolecular complexes. *Nature* **525**, 339-344, doi:10.1038/nature14877 (2015).

42 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).

43 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.



Firstly, 5 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 (See Methods): PEGs and CEGs. Secondarily, the functional TF networks of CEGs were predicted based on co-expression analysis and machine learning. And then the homologous and disease genes were annotated from existing databases. Finally, the functional significance of these TF networks was tested. Next, the hub TFs in the related TF networks were predicted based on the ARACNE algorithm. And then, CKO mice were constructed to verify the hub TFs and infer their GRN.

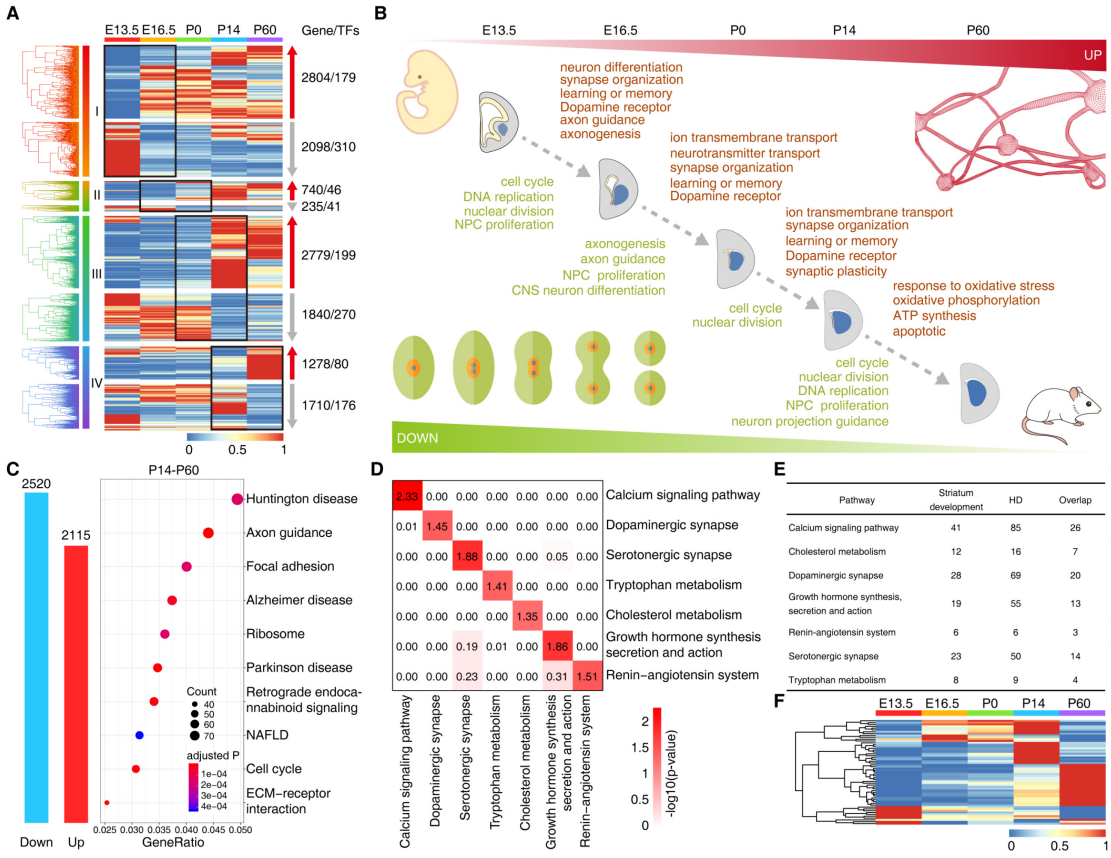


Fig. 2 Striatal development and correlation analysis with HD

A Expression patterns of DEGs between two adjacent time points. The number of genes and TFs are shown on the right.

B The patterning 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.

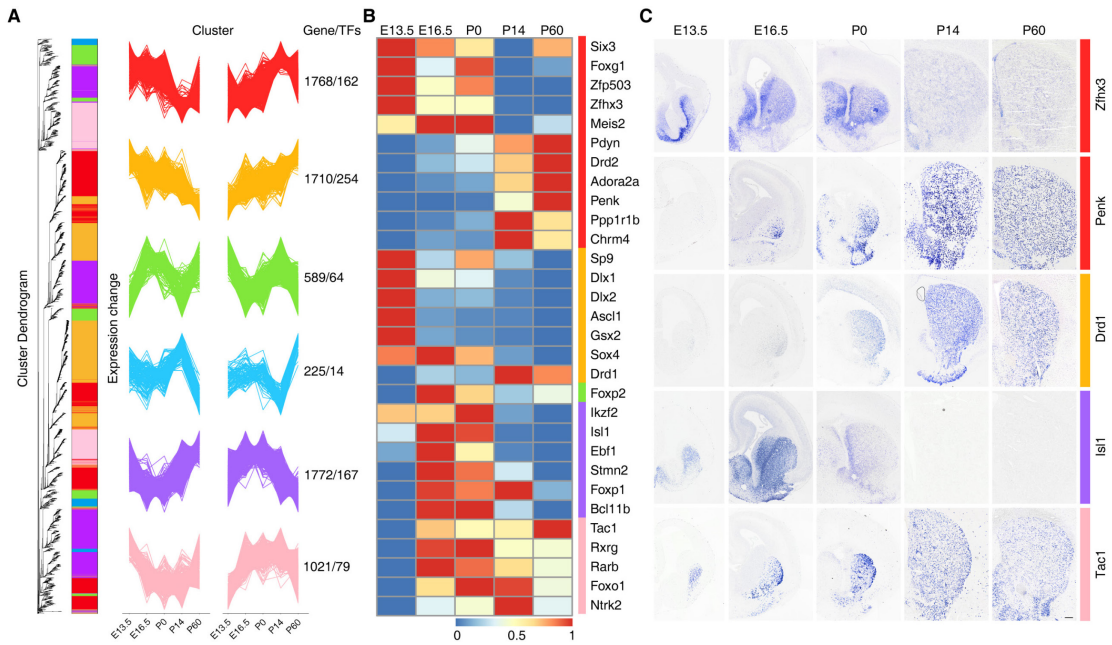


Fig. 3 Co-expression analysis of differentially expressed continuous genes

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. 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 in situ hybridization at 5 time-points. Scale bars: 100 μm in c.

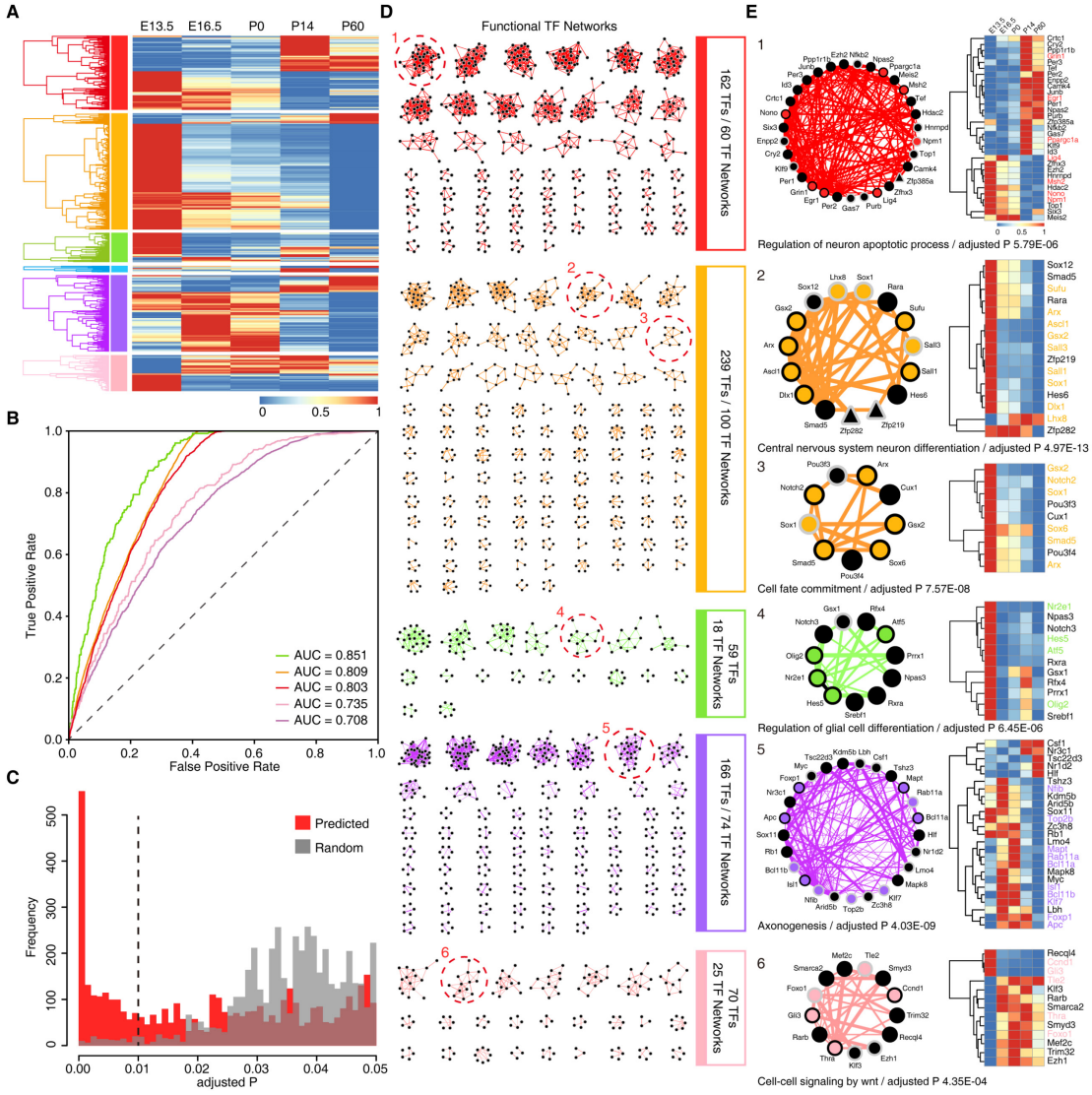


Fig. 4 Predicted potential functional TF networks based on machine learning

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 of 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.

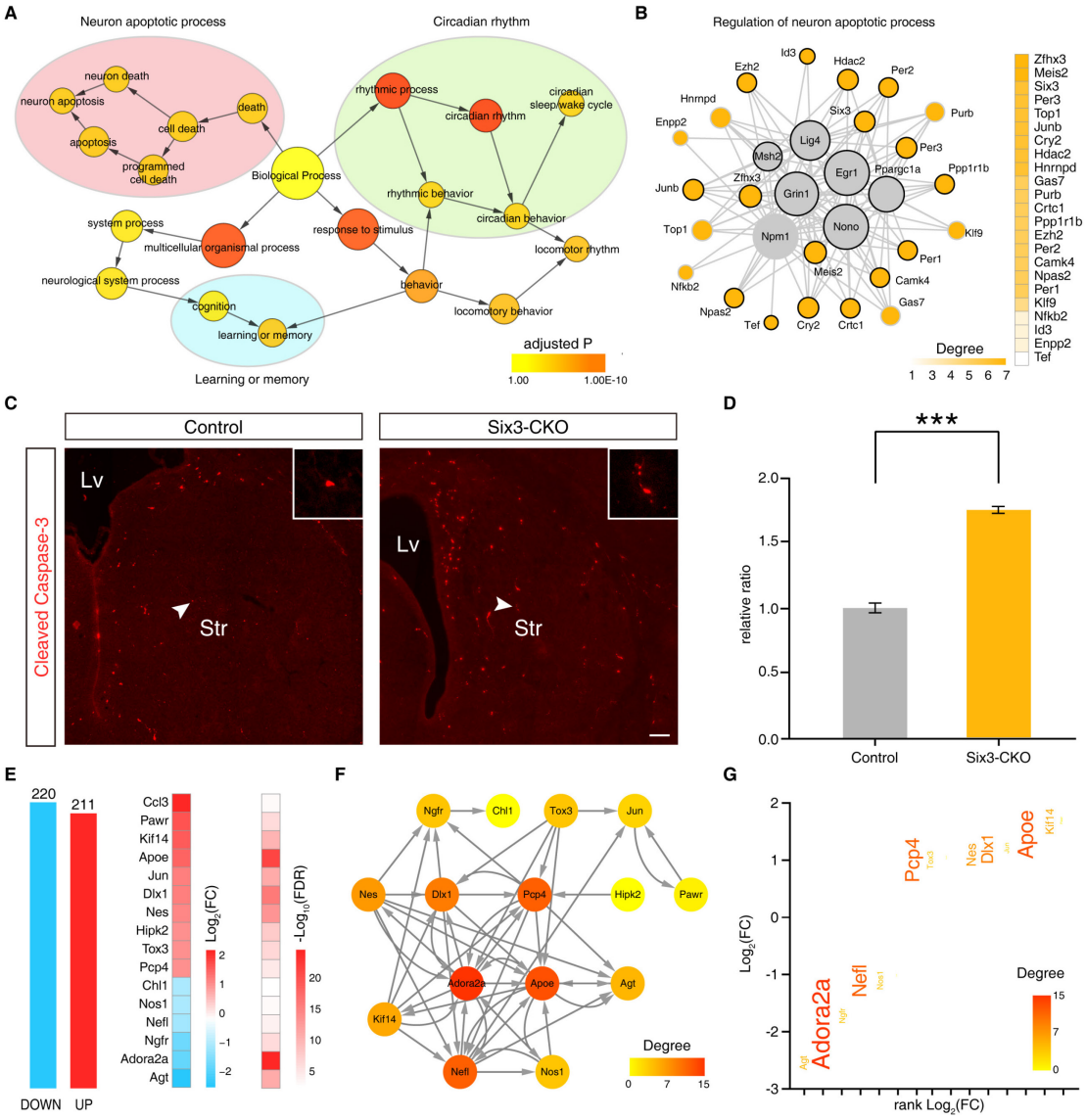


Fig. 5 Novel functions of Six3 predicted from TF networks

A Part of the BiNGO results 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

788 inferred interactions as edges. Grey nodes indicate the known functional TFs. Potential
789 functional TFs are coloured orange. The degree of orange nodes is shown on the right.
790 **C** A significant increase in expression of cleaved Caspase-3 in Six3-CKO striatum
791 compared to controls at P0. Scale bars: 100 μm in c.
792 **D** Histogram showing that the number of cleaved Caspase-3^+ cells was increased in
793 Six3-CKO mice. (Student's t-test, ***$P < 0.001$, n=3 mice per group, mean ± SEM). Lv:
794 lateral ventricle. Str: striatum.
795 **E** Histogram of DEGs in the LGE at P0 between control and Six3-CKO mice. The 16
796 DEGs associated with “regulation of neuron apoptotic process” are shown.
797 **F** A GRN of these “regulation of neuron apoptotic process” genes was inferred by
798 Genie3. The edge direction represents the regulatory interaction from regulatory genes
799 to target genes. Node size and colour are proportional to the number of genes (in-
800 degree and out-degree).
801 **G** Key downstream genes that are candidates for regulating the neuronal apoptotic
802 process. The significantly changed genes in Six3-CKO mice versus wild-type mice are
803 shown. Genes are coloured by their importance (degree) in the “regulation of neuron
804 apoptotic process” GRN, and the font size increases with the degree of node.
805
806
807
808
809

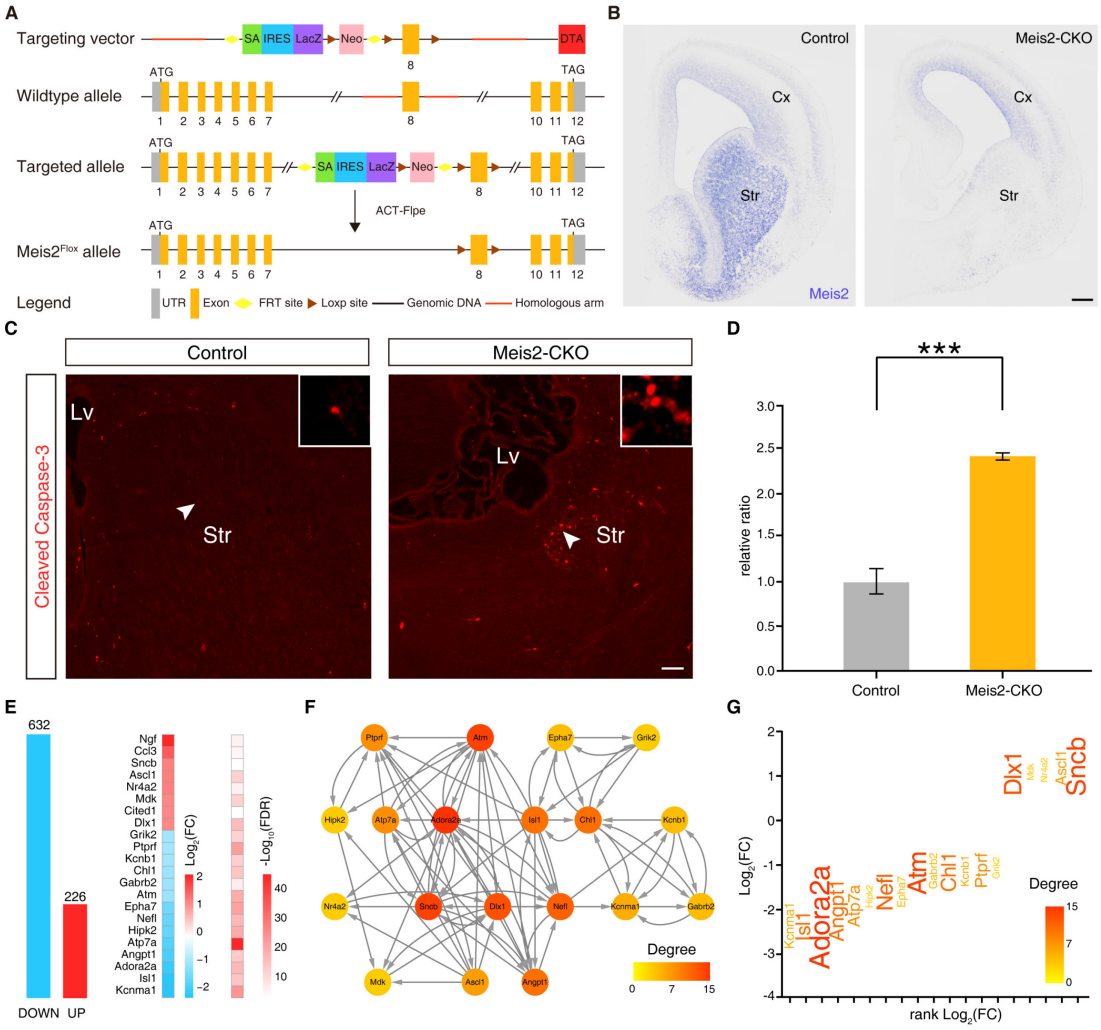


Fig. 6 Meis2 regulates neuronal apoptotic processes in the striatum

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 The immunofluorescence staining of cleaved Caspase-3 in Meis2-CKO striatum and 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.

820 **E** Histogram of DEGs in the LGE between control and Meis2-CKO mice at P0. The 18
821 DEGs associated with “regulation of neuron apoptotic process” are shown.
822 **F** A GRN of these “regulation of neuron apoptotic process” genes was inferred by
823 Genie3. The edge direction represents the regulatory interaction from regulatory genes
824 to target genes. Node size and colour are proportional to the number of genes (in-
825 degree and out-degree).
826 **G** The significantly changed genes in Meis2-CKO mice compared to control mice are
827 shown which are candidates for regulating the neuronal apoptotic process. Genes are
828 coloured by their importance (degree) in the “regulation of neuron apoptotic process”
829 GRN, and the font size increases with the degree of node.
830

831