Functional Identification of Specialized Basal Ganglia Circuits that Regulate Vocal Motor Sequences

Lei Xiao[1], Devin P. Merullo[1], Mou Cao[2], Marissa Co[1], Ashwinikumar Kulkarni[1], Genevieve Konopka[1], and Todd F. Roberts[1#]

[1]Department of Neuroscience UT Southwestern Medical Center, Dallas, TX, USA.

[2]Department of Pediatrics, UT Southwestern Medical Center, Dallas, TX, USA.

[#]Correspondence should be addressed to TFR (Todd.Roberts@utsouthwestern.edu)

SUMMARY

Production of learned vocalizations requires precise selection and accurate sequencing of appropriate vocal-motor actions. The basal ganglia are essential for the selection and sequencing of motor actions, but the cellular specializations and circuit mechanisms governing accurate sequencing of vocalizations are unknown. Here, we use single-nucleus RNA sequencing and genetic manipulations to map basal ganglia cell types and circuits involved in the production of songbird vocal sequences. We identify cell-type specializations in direct-like and indirect-like basal ganglia pathways, including evolutionary expansion of striatal and arkypallidal cell-types that could facilitate vocal sequencing. Surprisingly, we find that FoxP2, a gene important for vocal development, can potently and reversibly control accurate sequencing of adult birdsong, and that phasic dopamine selectively regulates repetition of syllables independent of its role in reinforcement-based learning of how they are sung. These findings identify key evolutionary specializations and circuits essential for selection and sequencing of vocal-motor actions necessary for vocal communication.

INTRODUCTION

Learning how to properly sequence motor actions is fundamental to coordinating movements and the development of new behavioral repertoires. The brain circuits that control sequencing of motor actions are thought to be hierarchically organized and span multiple levels of the brain (Hikosaka et al., 1999; Penhune and Steele, 2012; Wymbs et al., 2012; Yokoi and Diedrichsen, 2019). The basal ganglia direct and indirect pathways play integral roles in controlling motor behavior and the accurate production of motor sequences (Albin et al., 1995; DeLong, 1990; Jin and Costa, 2015). The core components of these circuits, including the cell-types and input and output pathways, are evolutionarily conserved among vertebrates (Grillner and Robertson, 2016; Stephenson-Jones et al., 2011) (Figure 1A). The emergence of new behavioral repertoires during evolution is hypothesized to involve duplication and expansion of basal ganglia 'modules' with similar direct and indirect pathways to support learning and control of the new behaviors (Grillner and Robertson, 2016; Grillner et al., 2013).

Although this basic idea likely holds for most coordinated movements, the evolution of highly specialized behaviors not broadly seen across taxa, such as learned vocal behaviors, have been associated with genetic and circuit specializations to these core or ancestral basal ganglia circuits. For example, human genetic disruptions of the transcription factor FOXP2, which is selectively enriched in direct-pathway medium spiny neurons (MSN)s, cause selective language development deficits marked by difficulties in controlling and accurately sequencing appropriate vocal-motor actions (childhood apraxia of speech), but do not disrupt the precise sequencing of non-vocal behaviors (Fisher et al., 2001; Lai et al., 2001; Vernes et al., 2011). Nonetheless, the cell types, input circuits, and output circuits involved in the selection and sequencing of appropriate vocal-motor actions are still poorly understood.

Zebra finches are perhaps the best studied vocal learning species, and like other songbirds, they have a specialized region (or 'module') in the dorsal striatum, termed Area X, that is dedicated to learning vocal-motor actions. Lesions of Area X in juvenile birds disrupt song learning and result in birds singing abnormal and highly variable vocal sequences from one rendition to the next (Bottjer et al., 1984; Scharff and Nottebohm, 1991; Sohrabji et al., 1990). Like the ancestral and mammalian striatum, Area X has MSNs that receive glutamatergic input from pallial/cortical regions and input from dopamine neurons in the substantia nigra and ventral tegmental area (VTA). These MSNs do not project outside of the striatum, but instead project onto pallidal-like neurons within Area X. These then provide inhibitory input to the thalamus, which is then relayed back to the pallium (Ding et al., 2003; Gale and Perkel, 2010; Person et al., 2008). Nevertheless, there are several functional and anatomical peculiarities of Area X that challenge simple one-to-one comparisons with ancestral or mammalian basal ganglia circuits and, ultimately, challenge our understanding of how this circuit regulates sequencing of vocal-motor actions.

The functional role of Area X in vocal-motor control is puzzling when compared to the generally understood role of the dorsal striatum in movement. The dorsal striatum and its dopaminergic inputs have an established role in developmental learning and in the regulation of movement initiation/vigor, sequencing, and termination (Coddington and Dudman, 2019; Jin and Costa, 2015; Klaus et al., 2019). The role of Area

X, however, appears to be more restricted to vocal development. Lesions of Area X or its dopaminergic inputs in adult birds do not have strong effects on song production(Bottjer et al., 1984; Hoffmann et al., 2016; Nordeen and Nordeen, 1993; Scharff and Nottebohm, 1991). Yet, expression of a mutant gene fragment that causes Huntington's disease in Area X of adult zebra finches leads to disruptions in sequencing of song syllables (Tanaka et al., 2016). Phasic increases or decreases in dopamine in Area X during singing also do not influence the vigor (amplitude), speed or structure of ongoing motor actions, as would be predicted based the role of phasic dopamine in the mammalian striatum (Xiao et al., 2018). Syllable-contingent phasic increases (Hisey et al., 2018; Xiao et al., 2018) and decreases (Xiao et al., 2018) in dopamine do, however, positively and negatively reinforce how song syllables are sung on future performances, implying a role in reinforcement learning that might be more akin to mammalian ventral striatal circuits.

In addition to these functional differences, several anatomical properties of Area X distinguish it from other portions of the avian and mammalian dorsal striatum (Figure 1C). First, MSNs in Area X corresponding to the canonical mammalian direct and indirect pathways MSNs have not been identified (Gale and Perkel, 2010). In traditional descriptions of mammalian circuits, the direct pathway MSNs are *Drd1*+ and *FoxP2*+ and project to the internal segment of the globus pallidus (GPi), while indirect pathway MSNs are *Drd2*+ and *FoxP2*- and project to the external segment (GPe) (Anderson et al., 2020; Gokce et al., 2016; Grillner and Robertson, 2016; Simonyan, 2019). Area X MSNs, however, have appeared to be a largely homogenous population expressing both *Drd1* and *Drd2* (Ding and Perkel, 2002; Kubikova et al., 2010). Also, unlike the ancestral and mammalian dorsal striatum, pallidal-like neurons are intermingled within Area X rather than being located in the pallidum (Carrillo and Doupe, 2004; Gale and Perkel, 2010; Person et al., 2008; Reiner et al., 2004).These Area X 'pallidal' neurons are also dissimilar from GPi and GPe neuron types because they lack direct projections to or from the subthalamic nucleus (STN) (Gale and Perkel, 2010). Together, these anatomical and functional peculiarities highlight potential gaps in our understanding of the cell-types and circuit specializations for accurate sequencing of vocalizations and how disruptions to these circuits may underlie speech disorders.

Here, we use single-nucleus RNA sequencing and genetic manipulations to map Area X cell-types and circuits involved in the production of songbird vocal sequences. We identify MSN cell-types corresponding to canonical direct and indirect pathways, and show that the pallidal-like cells in Area X appear similar to atypical arkypallidal GPe neurons which have been recently identified in mammals (Hegeman et al., 2016; Mallet et al., 2012). Using reversible genetic and optogenetic manipulations, we show that FoxP2 and phasic dopaminergic signaling in direct pathway MSNs play potent roles in regulating vocal sequences in adult birds. In addition, we find that disruptions in the control of vocal sequences can occur simultaneously to, but independent of, the reinforcement of the pitch of individual song syllables, suggesting a hierarchical representation of syntax (sequence selection) and sub-syntax action performance. Lastly, we find remarkable plasticity in the ability to recover appropriate vocal sequencing in adult birds, even following several months of genetic disruptions to the circuits that govern accurate sequencing of song. Together, these findings provide a new perspective on the organization and specializations of basal ganglia pathways

108 involved in the control of learned vocalizations, and provide insight into the potential circuit disruptions
109 that are associated with broad categories of speech and language disorders.
110

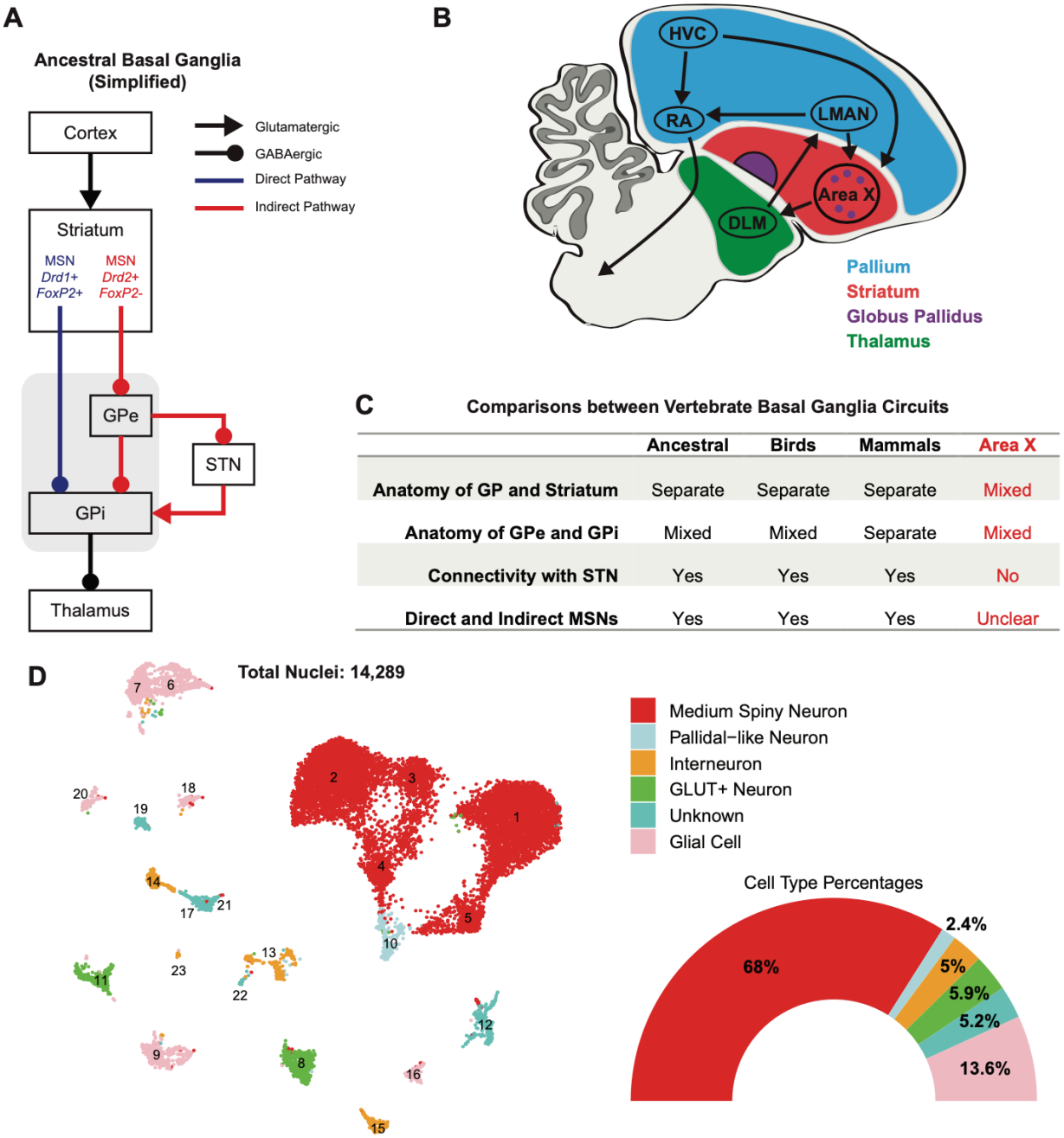
111 RESULTS

112 Single-Nuclei RNA Sequencing Reveals Taxonomy of Area X Cell Types

113 We set out to map the cell-types and circuits in Area X involved in the control and sequencing of learned
114 vocal motor actions. As a first step, we catalogued all the cell-types in Area X using snRNA-seq. We analyzed
115 14,289 nuclei pooled from Area X (4 hemispheres from 2 adult male zebra finches; Figures 1B, 1D, and S1).
116 A clustering analysis identified 23 distinct cell groups (Figure 1D). Functional identities were assigned to
117 each cell type based on established markers (Figure 1E). Most cells were MSNs (identified by the strong
118 collective expression of *Gad2*, *Ppp1r1b* (Reiner et al., 2004; Saunders et al., 2018), and *FoxP1* (Mendoza et
119 al., 2015)), with five MSN clusters comprising 68% of total cells. The pallidal-like cells, identified primarily by
120 the expression of *Penk* (Reiner et al., 2004) (see below for further discussion), exist within one cluster, and
121 are far less numerous at only 2.4% of all cells.

122 Aside from MSNs and pallidal-like cells, at least three major classes of striatal GABAergic interneurons were
123 present in Area X: *Pvalb*+ interneurons, *Sst*+/*Npy*+/*Nos1*+ interneurons, and *Chat*+ interneurons (Carrillo and
124 Doupe, 2004; Reiner, 2016; Reiner et al., 2004; Tepper and Bolam, 2004; Tepper et al., 2010). The *Pvalb*+
125 interneurons were split into two separate clusters, which may indicate two distinct *Pvalb*+ interneurons that
126 have been described previously in the striatum (Tepper and Bolam, 2004; Tepper et al., 2010) and also Area
127 X specifically (Reiner et al., 2004). Recently, a glutamatergic cell-type was reported to exist in Area X (Budzillo
128 et al., 2017). We identified two clusters expressing the glutamatergic transporter gene *Slc17a6*. The identities
129 of five clusters (12, 17, 19, 21, 22), comprising 5.2% of all cells, remain unknown based on gene expression
130 markers. Since the expression profiles of each of these clusters suggests a developmental origin from the
131 medial ganglionic eminence, these clusters may represent novel classes of interneurons, which are the only
132 cell type in the striatum known to originate from the medial ganglionic eminence (Chen et al., 2017). The
133 remaining clusters comprise various glial cell-types such as astrocytes and oligodendrocytes (Figure 1E).
134

135 **Figure 1 (next page). Overview of Basal Ganglia Circuits and Cell Types in Area X**
136 (A) Simplified schematic of basal ganglia circuits common across vertebrates. Not all connections are shown. The gray shaded
137 area indicates that in mammals the GPe and GPi are anatomically segregated, whereas in non-mammals they are mixed.
138 (B) Diagram of songbird neuroanatomy. Area X is a specialized striatal region that contains cells with pallidal-like properties.
139 (C) Comparisons between vertebrate basal ganglia circuits.
140 (D) A UMAP projection of nuclei from Area X, and a plot of cell types by percentage of total nuclei. Clusters are numbered in
141 ascending order by decreasing size (1-largest; 23-smallest).
142 (E) A heatmap of normalized expression for genes used to assign identities to cell types. Expression was normalized globally
143 across all genes, but a different scale is shown for each gene based on the highest normalized value.


A
Ancestral Basal Ganglia (Simplified)
Cortex
Striatum
MSN Drd1+ FoxP2+
MSN Drd2+ FoxP2-
GPe
STN
GPi
Thalamus
Glutamatergic
GABAergic
Direct Pathway
Indirect Pathway
B
HVC
RA
LMAN
Area X
DLM
Pallium
Striatum
Globus Pallidus
Thalamus
C
Comparisons between Vertebrate Basal Ganglia Circuits
Ancestral
Birds
Mammals
Area X
Anatomy of GP and Striatum
Separate
Separate
Separate
Mixed
Anatomy of GPe and GPi
Mixed
Mixed
Separate
Mixed
Connectivity with STN
Yes
Yes
Yes
No
Direct and Indirect MSNs
Yes
Yes
Yes
Unclear
D
Total Nuclei: 14,289
Medium Spiny Neuron
Pallidal−like Neuron
Interneuron
GLUT+ Neuron
Unknown
Glial Cell
Cell Type Percentages
68%
2.4%
5%
5.9%
5.2%
13.6%
E
1 (Medium Spiny Neuron)
5 (Medium Spiny Neuron)
2 (Medium Spiny Neuron)
3 (Medium Spiny Neuron)
4 (Medium Spiny Neuron)
10 (Pallidal−like Neuron)
14 (Pvalb+ Interneuron)
15 (Pvalb+ Interneuron)
13 (Sst+/Npy+/Nos1+ Interneuron)
23 (Chat+ Interneuron)
8 (GLUT+ Neuron)
11 (GLUT+ Neuron)
12 (Unknown Ppp1r1b+)
17 (Unknown GABA+)
19 (Unknown GABA+)
21 (Unknown GABA+)
22 (Unknown GABA+)
6 (Astrocyte)
7 (Astrocyte)
9 (Oligodendrocyte)
20 (Oligo. Precursor)
16 (Microglia)
18 (Endothelial)
Gad2
Slc17a6
Ppp1r1b
FoxP2
FoxP1
Tac1
Penk
Pvalb
Sst
Npy
Nos1
Chat
Lrig1
Mbp
Pdgfra
Csf1r
Flt1

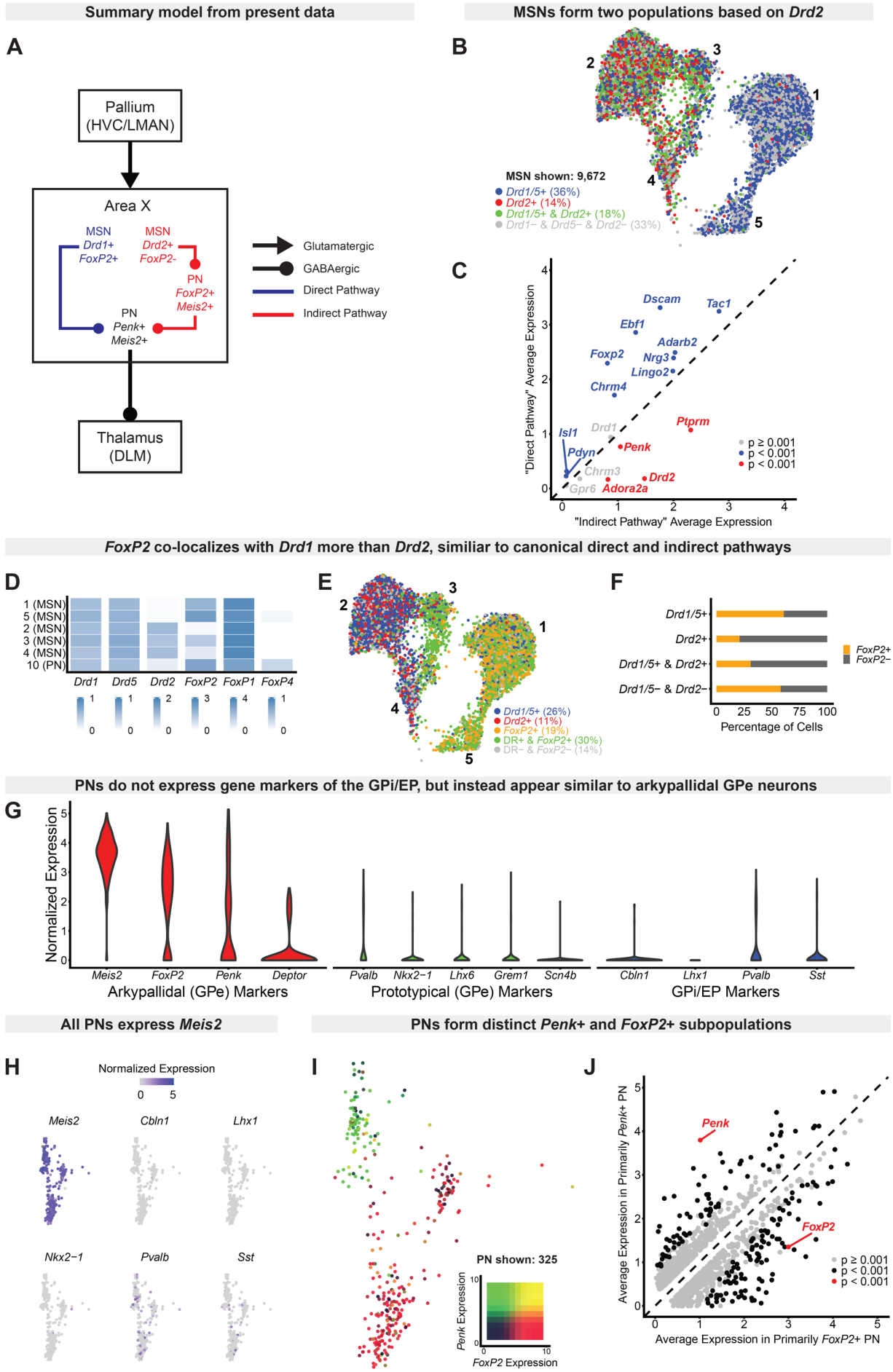
Identification of Area X Neurons Corresponding to Direct and Indirect Pathways

Area X MSNs corresponding to the direct and indirect pathways have not previously been identified (Gale and Perkel, 2010). Canonically, direct pathway MSNs are *Drd1*+ and *FoxP2*+, and project to the GPi while indirect pathway MSNs are *Drd2*+ and *FoxP2*-, and project to the GPe (Anderson et al., 2020; Gokce et al., 2016; Grillner and Robertson, 2016; Simonyan, 2019). Area X MSNs, on the other hand, have been thought to be a largely homogenous population expressing both *Drd1* and *Drd2,* and *FoxP2* (Ding and Perkel, 2002; Kubikova et al., 2010; Mendoza et al., 2015). It is unknown if MSNs selectively project to the thalamus-projecting and non-thalamus-projecting pallidal-like cells (Gale and Perkel, 2010).

Our snRNA-seq analysis revealed that 36% of the 9,672 MSNs are exclusively *Drd1*+ or *Drd5*+ (notated here as *Drd1/5+)*, while 13% are exclusively *Drd2*+. An additional 18% express varying levels of *Drd1/5* and *Drd2*. All five MSN clusters identified in our snRNA-seq dataset contain cells expressing *Drd1/5*, while only three contain cells expressing *Drd2* (Figure 2B). On the basis of this separation of *Drd2*+ cells across clusters, we performed a differential gene expression analysis by grouping the two clusters containing *Drd1*+/*Drd2*- cells into a putative "direct pathway," and the three clusters containing *Drd1*+ cells and *Drd2*+ cells (separate cells, not necessarily co-localizing *Drd1* and *Drd2*) into a putative "indirect pathway". The "direct pathway" grouping significantly expressed genes that classically mark direct pathway MSNs, including *FoxP2*, whereas the "indirect pathway" grouping significantly expressed genes that mark indirect pathway MSNs (Anderson et al., 2020; Gokce et al., 2016; Ho et al., 2018) (Figure 2C). Similar to what has been described in the mammalian direct and indirect pathways, *FoxP2* frequently co-localized with *Drd1*, but not with *Drd2* expressing neurons (Figures 2D and 2E). Of all cells that were *Drd1/5*+ and *Drd2*-, 61% also expressed *FoxP2*. However, of cells that were *Drd1/5*- and *Drd2*+, only 21% co-expressed *FoxP2* (Figure 2F).

Thus, MSNs in Area X can segregate into broad classes corresponding to the direct (*Drd1*+/*FoxP2*+) and indirect pathways (*Drd2*+/*FoxP2*-), as well as a third *Drd1*/2+ class. Further, these classes map to different clusters based on restrictive expression patterns of *Drd2* and *FoxP2* (Figures 2B-2F).

Figure 2 (next page). Identification of a Direct and Indirect Pathway Organization in Area X.
(A) A hypothesized model for Area X circuitry based on the present data.
(B) UMAP projection of MSN clusters, taken from Figure 1D. Nuclei are colored based on the exclusive expression of *Drd1/5*, *Drd2*, both, or neither. Percentages are as a total of MSNs.
(C) A differential gene expression analysis between nuclei grouped into a putative "direct pathway" (clusters 1 and 5) and "indirect pathway" (clusters 2, 3, and 4).
(D) A heatmap of normalized gene expression. Expression was normalized globally across all genes, but a different scale is shown for each gene based on the highest normalized value.
(E) UMAP projection of MSN clusters, taken from Figure 1D. Nuclei are colored based on the exclusive expression of *Drd1/5*, *Drd2*, *FoxP2*, any dopamine receptor and *FoxP2*, or none. Percentages are as a total of MSNs.
(F) A stacked bar plot illustrating the percentage of *FoxP2* co-expression in nuclei classified by dopamine receptor expression.
(G) A violin plot showing normalized expression of gene markers for arkypallidal GPe cells, prototypical GPe cells, and the GPi/EP.
(H) UMAP projection of PN clusters, taken from Figure 1D, with each nucleus colored by normalized expression of various genes.
(I) UMAP projection of PN clusters, taken from Figure 1D, with each nucleus colored according to the normalized expression of *FoxP2* and *Penk*. Expression value colors are relative to each gene (0-lowest value for that gene; 10-highest value for that gene).
(J) A differential gene expression analysis between PN that are primarily *FoxP2* and PN that are primarily *Penk*, identified from a sub-cluster analysis.


Summary model from present data
MSNs form two populations based on Drd2
FoxP2 co-localizes with Drd1 more than Drd2, similiar to canonical direct and indirect pathways
PNs do not express gene markers of the GPi/EP, but instead appear similar to arkypallidal GPe neurons
All PNs express Meis2
PNs form distinct Penk+ and FoxP2+ subpopulations

Area X Pallidal Neurons Resemble Specialized GPe Arkypallidal Neurons

We next investigated the molecular makeup of pallidal cells in Area X. Two morphologically and physiologically distinct pallidal-like cells have been described in Area X (Farries et al., 2005a). The cell-type that projects out of Area X to the dorsal thalamic nucleus DLM (medial portion of the dorsolateral thalamus) is larger and displays high-frequency (>60Hz) continuous discharge of action potentials, whereas the cell-type that does not project outside Area X is smaller and exhibits high-frequency firing rates characterized by bursting and intermittent pauses in activity (Goldberg et al., 2010). Pallidal-like cells in Area X also have unique anatomical properties, such as being anatomically intermingled with striatal cells, having putative projections back onto MSNs, and a lack of projections to the STN or substantia nigra pars reticulata (SNr) (Gale and Perkel, 2010). Although very little is known about their molecular properties, it is known that the pallidal-like cells in Area X immunolabel for enkephalin, LANT6 (a neurotensin-related peptide distinct from neurotensin and neuromedin N (Reiner, 1987; Reiner and Carraway, 1987)), and do not immunolabel for the protein product of *Nkx2-1*, which typically marks cells of pallidal origin (Carrillo and Doupe, 2004; Reiner et al., 2004).

We identify here Cluster 10 as containing the pallidal-like cells because of its strong expression of *Penk* (the gene that encodes the precursor for enkephalin), its expression of other pallidal neuron markers identified in mammals (see below), and the lack of other clusters with an expression profile consistent with what is known about the pallidal-like cells. For example, other clusters expressing *Penk* (Clusters 15, 13, and 22) are likely interneurons due the co-expression of markers for striatal interneurons suggesting an origin from the MGE, such as *Nkx2-1*, *Lhx6*, Sox6 (Chen et al., 2017). While cluster 10 also expresses *Ppp1r1b*, which classically marks MSNs (Reiner et al., 2004; Saunders et al., 2018), *Ppp1r1b* is also expressed in Area X glial cells, indicating that it alone is not an exclusive marker of avian MSNs at the level of immature nuclear mRNA. We did not detect expression of *Nkx2-1* or *Nts* (the gene encoding the protein precursor to LANT6) in Cluster 10. Compared to all other clusters in our dataset, cluster 10 is the most consistent with pallidal-like cells, but we cannot exclude the possibility that other clusters may also contain pallidal-like cells.

We next asked if Area X pallidal neurons have molecular profiles similar to those described in the mammalian GPi and GPe. Contrary to our expectations, gene markers for the GPi (*Cbln*, *Lhx1*, *Pvalb*, and *Sst*) (Saunders et al., 2018; Wallace et al., 2017) are largely absent from Area X PNs (i.e. cluster 10) and also do not co-express within any other cluster in our dataset (Figures 2G, 2H and, S2). This suggests that the pallidal-like cells are not migrations of cells from the GPi. The GPe contains two major pallidal cell-types termed "prototypical", which project to the STN, and "arkypallidal", which do not project to the STN but instead project back onto MSNs in the striatum (Abdi et al., 2015; Mallet et al., 2012). We found that markers for the prototypical cells (*Pvalb*, *Nkx2-1*, *Lhx6*, *Grem1*, and *Scn4b*) are also generally absent from the pallidal-like cells in cluster 10 or other clusters in our dataset (Figures 2G and 2H). Rather, the cluster containing pallidal-like cells expresses high levels of *Meis2*, *FoxP2*, *Penk*, and *Deptor*, which are all prominent markers of arkypallidal cells (Figures 2G and 2H). It is significant to note that *Meis2* is a developmental marker of the lateral ganglionic eminence, from which arkypallidal cells and striatal cells emerge, and that *Nkx2-1* is a marker of the medial ganglionic eminence, from which the pallidum and prototypical GPe cells derive (Nóbrega-Pereira et al., 2010) (Figures 2G and 2H).

228 Unlike arkypallidal cells identified in mammals, which co-express *FoxP2* and *Penk*, Area X pallidal neurons do
229 not co-express these genes (Figure 2I). Rather, a sub-clustering analysis revealed that although *Penk* and
230 *FoxP2* are each strongly expressed, this occurs in non-overlapping populations of neurons (Figures 2J and S2B
231 and S2C). Based on previous findings indicating that thalamus-projecting Area X pallidal cells express
232 enkephalin, we propose that the *Penk+* pallidal-like cells correspond to the thalamus-projecting cells
233 (previously described as GPi-like), and that the *FoxP2+* cells are the non-thalamus-projecting cells (previously
234 described as GPe-like) (Farries et al., 2005a; Goldberg et al., 2010).

235 Together, our transcriptome analysis indicates that pallidal cells in Area X, at least in part, molecularly
236 resemble arkypallidal cells of the GPe. This suggests that arkypallidal cells were likely already present in stem
237 amniotes and that evolution of basal ganglia circuits for motor control of learned vocalizations involved
238 diversification of arkypallidal cells into two separate classes, one of which may project to the dorsal thalamus.
239 Moreover, this finding provides an evolutionary context in which to consider the lack of an Area X projection
240 to the avian homologue of the STN and the previously described projection from Area X pallidal neurons back
241 onto MSNs, both core hodological features of mammalian arkypallidal neurons. When considered with our
242 broader description of cell-types, we suggest that Area X contains at least three populations of MSNs, two of
243 which correspond closely to canonically described direct and indirect pathways. These cell clusters are best
244 delineated by differential expression of *FoxP2* and *Drd2*. Although more detailed mapping of synaptic
245 connections in this circuit is needed, we hypothesize that the direct-like *Drd1*+/*FoxP2*+ MSNs project to the
246 *Penk*+ arkypallidal neurons, which then project to the thalamic song-control region DLM, and that the
247 indirect-like *Drd2*+/*FoxP2*- MSNs project to the non-thalamic projecting *FoxP2*+ arkypallidal neurons (Figure
248 2A).

249

250 Knockdown of FoxP2 Decreases Dopamine Receptor Expression in the Direct Pathway and Causes 251 Progressive Disruptions in Syllable Sequencing

252 The role of Area X in the selection and sequencing of song elements is largely unknown. As mentioned
253 previously, lesions of Area X in adult birds have only a limited impact on song. Our mapping of Area X cell-
254 types and their relation to components of the direct and indirect pathways, including components previously
255 implicated in regulating motor sequences in rodents, spurred us to take a fresh look at the role of adult Area
256 X in song motor control. We first focused on the role of FoxP2 in regulating adult song. Knockdown of FoxP2
257 impairs song learning in juvenile birds, but is not known to impair adult song (Figure 3A). However, previous
258 studies into the role of FoxP2 in adult song production have expressed shRNA using vesicular stomatitis virus
259 glycoprotein (VSV-G) pseudotyped lentivirus, which only sparsely infects avian neurons(Roberts et al., 2010).

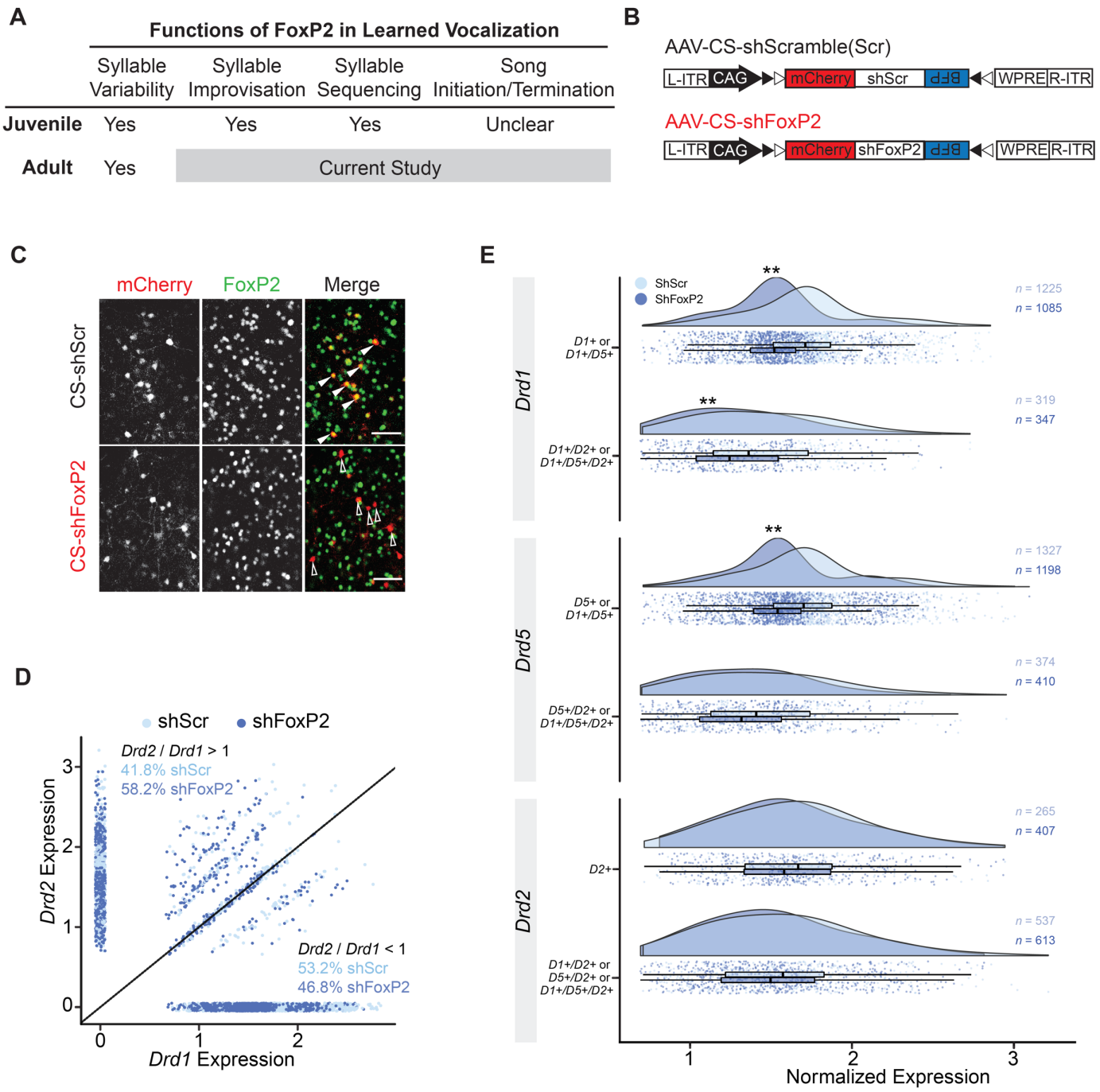
260 To more stringently test the potential role of FoxP2 in Area X we sought an approach that drove broad and
261 genetically reversible knockdown of FoxP2 in zebra finch neurons. We employed a Cre-switch (CS) AAV
262 platform, which strongly expresses in finch neurons (Figure 3B). In our constructs, shRNA is constitutively
263 driven by the Pol II promoter CAG, and expression of the shRNA can be turned-off with introduction of Cre
264 recombinase(Saunders et al., 2012; Yu et al., 2015). We tested CS-shFoxP2 and CS-shScramble (CS-shScr)
265 constructs in vitro and in vivo and found that we could significantly reduce the expression of FoxP2 following

266 expression of CS-shFoxP2, and that a subsequent injection with a second AAV expressing Cre-GFP could
267 rescue the knockdown of FoxP2 (Figures 3B, C and S4).

268 Increased dopamine levels and decreased dopamine receptor expression have been associated with
269 impairments in coordinated movements(Chen et al., 2013; Girasole et al., 2018), and disruptions of FoxP2
270 expression have been shown to result in increased dopamine levels and reduced expression of dopamine
271 receptors in the striatum (Enard et al., 2009; Murugan et al., 2013). Therefore, we used snRNA-seq to directly
272 measure the influence of FoxP2 knockdown on expression of dopamine receptor transcripts in the direct and
273 indirect pathway MSNs identified above. We measured nuclear expression levels of *Drd1*, *Drd2,* and *Drd5*
274 transcripts in adult male zebra finches injected with either CS-shFoxP2 or CS-shScr (transcriptomes from an
275 additional 12,956 Area X cells collected from 4 hemispheres (n = 2 birds) injected with CS-shFoxP2 (Figures
276 S3A and S3B)). Analysis of 5,845 MSNs co-expressing *FoxP2* and any dopamine receptor (*Drd1*/*Drd5*/*Drd2*)
277 revealed that knockdown of FoxP2 causes an overall decrease in the *Drd1* to *Drd2* ratio across MSNs in Area
278 X. This change in the *Drd1* to *Drd2* ratio resulted from decreased expression of *Drd1* and *Drd5* in the direct
279 pathway MSNs (Figures 3D and 3E).

280

281 **Figure 3 (next page). A Cre-switch Platform for Exploring the Function of FoxP2 in Learned Vocalizations.**
282 (A) Summary table. The functions of FoxP2 in learned vocalization, including syllable variability, improvisation, sequencing
283 and song initiation and termination in either juvenile(Haesler et al., 2007; Murugan et al., 2013; Norton et al., 2019) or adult
284 birds(Day et al., 2019; Murugan et al., 2013).
285 (B) AAV constructs to achieve Cre-dependent silencing of zebra finch FoxP2. A single transgene containing open reading
286 frames of two fluorophores (mCherry and BFP) oriented in opposite directions switches between expressing mCherry or BFP
287 depending on Cre-driven recombination. Small hairpin RNAs against zebra finch FoxP2 gene(shFoxP2) as well as scrambled
288 hairpin (shScr) were inserted into the 3'UTR of the mCherry in the functional orientation with respect to the CAG promoter.
289 For this Cre-Switch (CS) configuration, the expression of mCherry and shRNAs are only maintained in the absence of Cre,
290 whereas the expression of BFP is activated in the presence of Cre. WPRE, woodchuck polyresponse element. ITR, inverted
291 terminal repeats. Open and filled triangles indicate loxP and lox2272, respectively.
292 (C) FoxP2 expression in the cells infected with CS-shScr (top, filled triangles) and CS-shFoxP2 (bottom, open triangles) in Area
293 X of adult birds. mCherry positive cells are cells infected with CS constructs. Scale bar, 50 µm.
294 (D) A scatterplot illustrating the correlation between *Drd1* normalized expression and *Drd2* expression in *FoxP2*+ striatal cells
295 that are also *Drd1/5*+ or *Drd2*+. Each point represents a cell. Dashed line indicates the line of equality. Cells above the line
296 have a ratio of *Drd2*/*Drd1* greater than 1. Cells below the line have a ratio of *Drd2*/*Drd1* below 1.
297 (E) Distribution and boxplots of normalized expression for *Drd1*, *Drd5*, and *Drd2* in *FoxP2*+ MSNs in CS-shScr+ and CS-
298 shFoxP2+ birds. Cells are grouped by the expression of dopamine receptors, e.g. *D1+* indicates the cells express *Drd1+* only
299 and no other dopamine receptor, whereas *D1*+/*D5*+ indicates the cells express both *Drd1* and *Drd5*. The expression level of
300 *Drd1* in cells from CS-shFoxP2+ birds was significantly lower than in CS-shScr+ birds ($p < 0.001$, Welch's t-test) in all *Drd1*+
301 cells. The expression of *Drd2* in *FoxP2*+ MSNs did not differ between the two groups. The lower and upper bounds of the
302 boxes indicate the 25th and 75th percentiles; the whiskers extend in either direction from the bound to the furthest value
303 within 1.5 times the interquartile range.



To test the effect of FoxP2 knockdown on vocal behavior, we made bilateral injections of CS-shFoxP2 into Area X of adult birds (159 ± 18 dph) and monitored their song over many weeks. Adult zebra finch song is normally highly stereotyped and characterized by precise ordering of individual song syllables, but we found that knockdown of FoxP2 resulted in disruptions in the overall song syntax, affecting the sequencing of song syllables within three weeks of viral injections (Figures 4E, 4F, 7C and S5A). Overall, changes to song included anomalous repetition of individual syllables, replacement and deletion of some syllables from the song, distortion of existing song syllables, and the creation and insertion of entirely new song syllables. When singing, birds appeared to get caught in motor loops, repeating a certain song syllable many times before transitioning to the next syllable. We found that gross changes in song syntax became progressively more

severe and persisted for as long as birds were recorded (2-6 months; Figures 4E, 4F, 7C and S5A). Although the severity of song syntax disruptions varied from bird to bird, increased repetition of syllables at the beginning and/or end of song motifs was observed in all birds, suggesting that FoxP2+ neurons in Area X contribute to the regulation of the initiation and termination of vocal sequences and the transition from one syllable to the next.

Although we observed disruptions in song syntax, we failed to detect significant changes in variability of fundamental frequency or entropy variance of song syllables following FoxP2 knockdown (Figures 4C and 4D). This indicates that syntax changes did not emerge from a general lack of motor control. Rather, it suggests that syntax disruptions likely result from maladaptive changes to higher-order representations involved in the selection and sequencing of vocal motor actions.

We tested if the observed changes in song depended on the knockdown of FoxP2 by injecting separate groups of birds with an AAV expressing either our scrambled construct (CS-shScr) or our knockdown construct (CS-shFoxP2) combined with one expressing Cre-GFP (inverted Cre-switch, ICS). We did not observe changes in song structure or an increase in syllable repetitions in any of these birds (Figures 4A and 4B). Together, these results demonstrate an essential and previously unappreciated role for FoxP2 in the selection and sequencing of vocal-motor actions and suggest that dopamine signaling, mediated by Drd1/5, could also be involved.

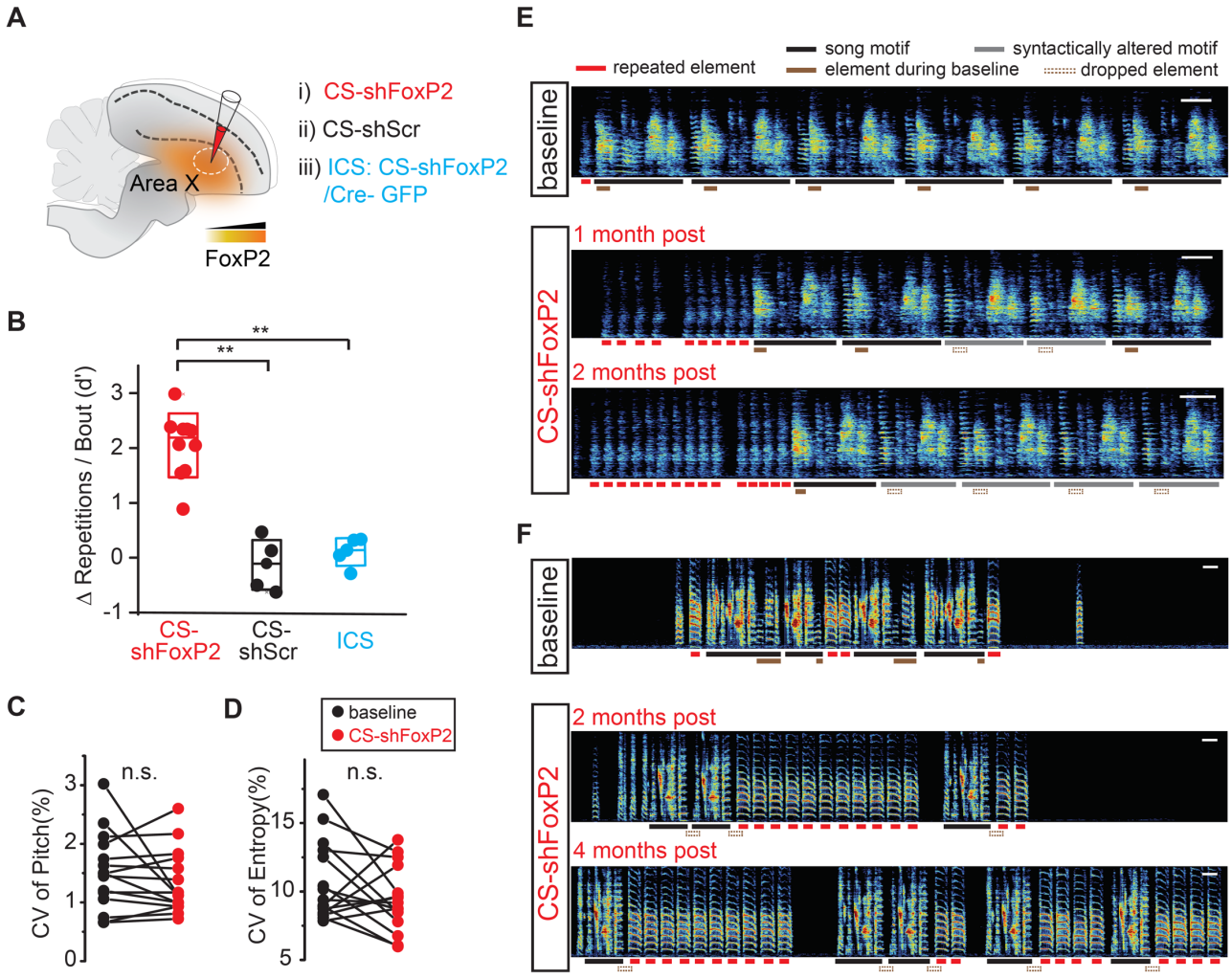


Figure 4. Diminished FoxP2 Level in Area X Caused Disruptions in Syllable Sequencing in Adulthood.
(A) CS constructs were bilaterally injected into Area X of adult birds, alone (i, CS-shFoxP2; ii, CS-shScr) or with Cre-GFP (iii, CS-shFoxP2/Cre-GFP, termed Inverted Cre Switch (ICS)).
(B) Changes in the number of vocal repeats per song bout, expressed in units of *d'*, for CS-shFoxP2+ birds(red circles, *d'*=2.05±0.41, n=10 syllables from 8 birds), CS-shScr+ birds (black circles, *d'*= -0.13±0.56, n = 5 syllables from 5 birds), and ICS+ birds (blue circles, *d'*=0.11±0.31, n = 5 syllables from 5 birds). Changes in the number of repetitions of vocal elements in CS-ShFoxP2+ birds were significantly greater than changes observed in CS-shScr+ and ICS+ birds (CS-shScr, p= 0.0023; ICS, p=0.015, Kruskal-Wallis test). Box indicates the median ± 1.0 SD, mean shown by open dot.
(C) Variability in pitch of syllables for baseline day (black, coefficient of variation [CV] = 1.52% ±0.17%) and two months post injection of CS-shFoxP2+ birds (red, CV = 1.34% ±0.14%). Knockdown of FoxP2 in Area X did not change the coefficient of variation of pitch (p = 0.75, n = 15, Wilcoxon matched-pairs signed-rank test).
(D) Variability in entropy of syllables for baseline day (black, CV = 10.7% ±0. 75%) and two months post injection of CS-shFoxP2+ birds (red, CV = 9.46% ±0.62%). Knockdown of FoxP2 in Area X did not change the coefficient of variation of pitch (p = 0.17, n = 15, Wilcoxon matched-pairs signed-rank test).
(E) Spectrograms of song recorded on the baseline day, 1 month, and 2 months after bilateral injection of CS-shFoxP2 construct in Area X of an adult bird. The number of repetitions of introductory elements in each song bout were increased over time, and one syllable was omitted over the course of 2 months. Song bouts consist of a series of introductory elements followed by a sequence of syllables produced in a stereotyped order, referred to as the 'motif'. Scale bar, 200 ms.
(F) Spectrograms of song recorded from a second bird on the baseline day, 2 months, and 4 months after bilaterally injection of CS-shFoxP2 construct in Area X. The number of repetitions of one syllable gradually increased over time, whereas other vocal elements were omitted over the course of 4 months. Scale bar, 200 ms.

Phasic Dopamine in Area X Drives Maladaptive Syllable Repetitions Independent of Reinforcement Learning

Following these observations, we were interested in testing whether dopamine signaling by itself plays a role in the regulation of vocal-motor sequences. Recent studies have emphasized the idiosyncratic role of dopamine in reward-based learning and the control of coordinated movements(Coddington and Dudman, 2019). We have previously shown that acute manipulation of dopaminergic inputs to Area X does not affect ongoing vocal-motor actions but can instruct rapid bidirectional learned changes in song, consistent with reinforcement-based models for motor learning (Xiao et al., 2018). More broadly however, it is known that physical manifestations of dopamine related impairments, like chorea or akinesia, can progressively develop with long term disruptions in dopamine signaling(Chen et al., 2013). Thus, we wondered if sustained manipulations of dopaminergic tone or phasic signaling would lead to disruptions in the sequencing of vocal-motor actions. To test these ideas, we directly manipulated dopaminergic circuits across several days in freely singing birds using pharmacological and optogenetic approaches.

First, we tested whether chronic elevation of dopamine could elicit gross changes in song. We implanted reverse microdialysis probes bilaterally in Area X and individually infused dopamine, Drd1-like agonist, and Drd2-like agonists, each for several days (Figures S6A and S6B). Remarkably, and in contrast to the vocal deficits observed following FoxP2 knockdown, birds continued to sing at normal rates and without disruptions in song spectral structure or syntax during direct infusion of dopamine or dopamine receptor agonists (Figure 6C).

Next, we tested if sustained manipulations to phasic dopamine activity would result in disruptions in song sequences. Direct pathway MSNs, which express *FoxP2*, are thought to be sensitive to phasic increases in dopamine, while indirect pathway MSNs are thought to be more sensitive to phasic decreases in dopamine

(Dreyer et al., 2010; Richfield et al., 1989). Therefore, we systematically tested the effect of both phasic increases and phasic decreases in dopamine release over several days. An adeno-associated virus (AAV) expressing an axon-targeted channelrhodopsin (ChR2), archaerhodopsin (ArchT), or green fluorescent protein (GFP) was injected bilaterally into VTA, and birds were implanted with optical fibers over Area X (Figures S7A and S7B). Bilateral optical illumination of VTA axon terminals in Area X was targeted to an individual syllable in the song motif for 3-12 consecutive days (Figure 5A). Light pulse delivery was dependent on natural trial-to-trial variation in the pitch of the targeted syllable. In agreement with our previous results (Xiao et al., 2018), we found that optogenetic excitation and inhibition elicited learned changes in the pitch of the targeted syllable on future performances (Figure 5C). In addition to changes in the pitch of the targeted syllable, we found that phasic increases in dopamine also resulted in disruptions in song syntax similar to birds in which we knocked down FoxP2 expression in Area X (Figures 5D, 5E and S7C). Phasic excitation of dopaminergic terminals in Area X resulted in a significant increase in the number of times birds repeated syllables at the beginning and/or end of their song motif (Figures 5B, 5D and 5E). This increase in syllable repetitions was observed in all ChR2 expressing birds by the third day of phasic stimulation. We did not observe disruptions in the selection and sequencing of vocal motor actions in birds that received phasic inhibition of dopamine release during singing, or in birds expressing GFP (Figures 6B, 6C , S7D and S7E).

Figure 5 (next page). Song Contingent Excitation of Dopamine Terminals in Area X Induces Dysfluent Repetition of Song Syllables.

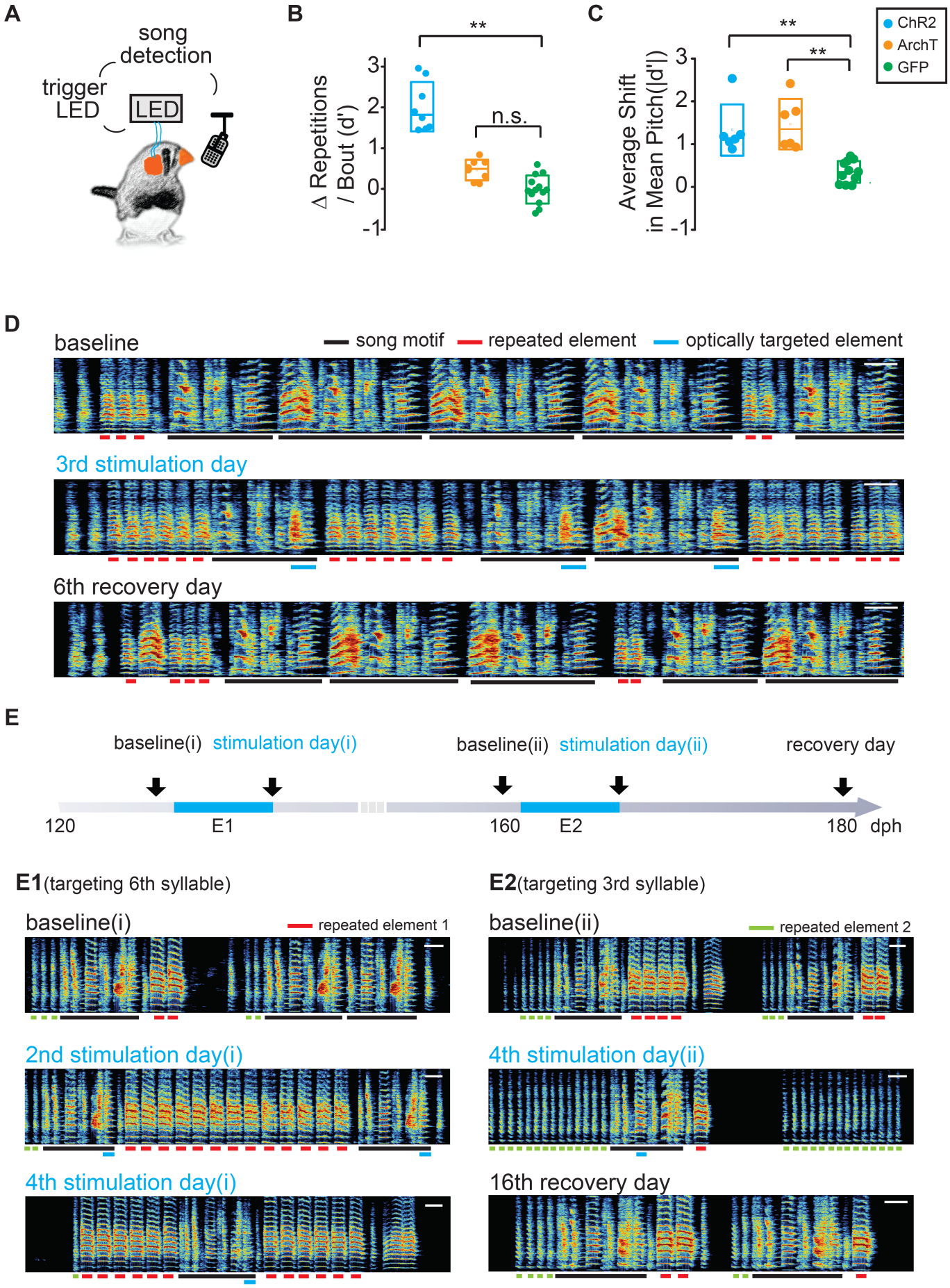
(A) Schematic of closed loop optogenetic experimental paradigm.

(B) Changes in the number of repetitions of vocal element per song bout between the baseline day and last illumination day, expressed in units of *d'*, for ChR2+ birds (blue circles, *d'*=2.02±0.5, n=8 vocal elements from 6 birds), ArchT+ birds (orange circles, *d'*=0.46±0.21, n = 8 syllables from 6 birds), and GFP+ birds (green circles, *d'*=-0.015±0.2, n = 13 syllables from 4 birds). Change in the number of repetitions of vocal elements in ChR2+ birds were significantly greater than change in GFP+ birds ($p<0.0001$, Kruskal-Wallis test), but there was no significant difference between ArchT+ and GFP+ birds ($p=0.12$, Kruskal-Wallis test). Box indicates the median ± 1.0 SD, mean shown by open dot.

(C) Average shift in mean pitch, expressed in units of $|d'|$, for ChR2+ birds (blue circles, $|d'|$ =1.33±0.6, n=6 syllables from 6 birds), ArchT+ birds (orange circles, $|d'|$ =1.47±0.59, n=6 syllables from 6 birds), and GFP birds (green circles, $|d'|$ =0.35 , n = 11 syllables from 6 birds). Average shift in mean pitch for both ChR2+ and ArchT+ birds were higher than 0.75, and also significantly higher than control GFP+ birds (ChR2+, $p=0.0025$; ArchT+, $p=0.0025$; Kruskal-Wallis test). Box indicates the median ± 1.0 SD, mean shown by open dot.

(D) Spectrograms of song recorded from a ChR2+ bird on the baseline day, 3rd stimulation day and 6th recovery day. Light pulses (~455 nm, 100ms) were delivered over the target syllable during lower pitch variants but not during higher pitch variants.

(E) Schematic of experiment in a ChR2+ bird in which light pulses (~455 nm, 100ms) were delivered over two different target syllables at different times over the course of two months. The bird starts repeating a song element either at the end of its motif (E1) or the beginning of its motif (E2) or both (E1 bottom), depending on which syllable in the song was optogenetically targeted. Scale bar, 200 ms.


A
song detection
trigger LED
LED
B
Δ Repetitions / Bout (d')
**
n.s.
C
Average Shift in Mean Pitch(|d'|)
**
**
ChR2
ArchT
GFP
D
baseline
song motif
repeated element
optically targeted element
3rd stimulation day
6th recovery day
E
baseline(i)
stimulation day(i)
baseline(ii)
stimulation day(ii)
recovery day
120
E1
160
E2
180
dph
E1(targeting 6th syllable)
baseline(i)
repeated element 1
2nd stimulation day(i)
4th stimulation day(i)
E2(targeting 3rd syllable)
baseline(ii)
repeated element 2
4th stimulation day(ii)
16th recovery day

414

415 The increased repetition of song syllables generalized to all song performances, not just optically stimulated
416 trials, and persisted for 2 or more days after optical stimulations were discontinued (Figures 6D and 6E). This
417 suggests that phasic increases in dopamine do not have a direct influence on ongoing vocal-motor actions
418 but may maladaptively influence higher-order representations involved in the selection and sequencing of
419 vocal motor actions. Consistent with this interpretation, in only one case (1 of 8 birds) did the bird start to
420 repeat the optically targeted syllable (Figure S7C). In this case the bird began to repeat a pair of syllables in
421 the middle of its song motif, suggesting that vocal repetitions could emerge at any point in song and are not
422 necessarily confined to the beginning and ending of the song motifs. Also consistent with a maladaptive
423 influence on higher-order representations, we found that the syllable which was optically stimulated
424 influenced which other syllable(s) birds began to repeat (Figure 5E). In a bird with a six-syllable song,
425 optogenetic excitation of the sixth syllable (Figure 5E1) resulted in vocal repetitions at termination of the
426 song motif, while optogenetic excitation of the third syllable (Figure 5E2) resulted in vocal repetitions at the
427 initiation of the song motif.

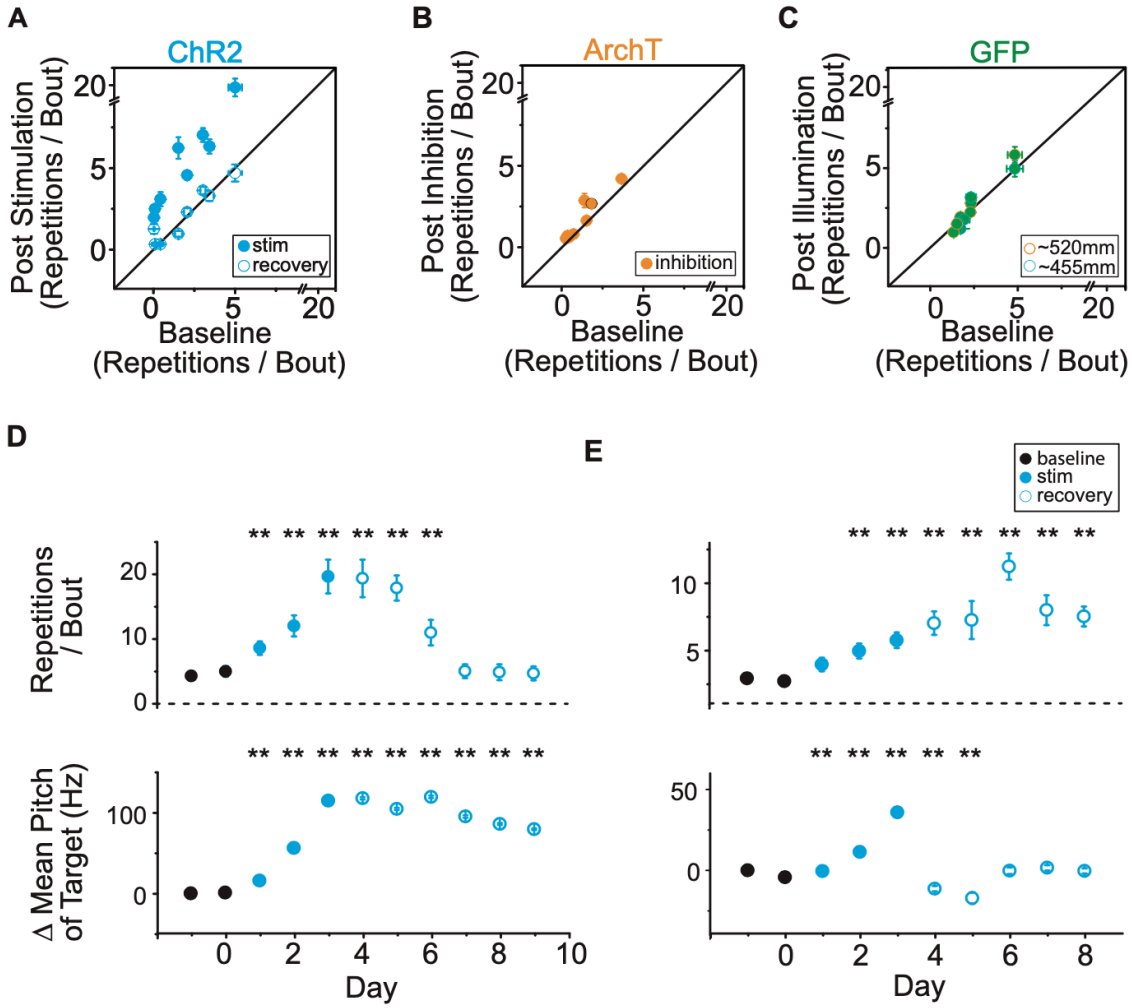
428 Notably, optogenetic stimulation of dopamine terminals did not recapitulate the full range of vocal
429 disruptions observed following FoxP2 knockdown. For example, we did not observe dropping of syllables or
430 creation of new syllables following optogenetic stimulation, but it did cause significantly increased rates of
431 vocal repetition, which was the most consistent effect of FoxP2 knockdown. We suspected that knockdown
432 of FoxP2 could have broad effects on gene expression, which in turn may account for the broad, knockdown-
433 related, disruptions in song. We used the ToppGene database to perform a gene ontology analysis on the
434 3,388 genes that were significantly different between the *FOXP2+* cells in the scramble and knockdown
435 groups, after adjusting for multiple comparisons. Many different families of gene function showed decreased
436 gene expression in the knockdown group, such as those involved in cell-cell signaling, neurogenesis, neuron
437 projections, and synapses (Figure S3C-D). Together, these results suggest that FoxP2 expression in striatal
438 circuits can influence the production and control of vocalizations in multiple ways and that relationship
439 between phasic dopaminergic signaling and FoxP2 may more selectively disrupt precise sequencing of
440 vocalizations and lead to maladaptive repetition of song syllables.

441 Because disruptions in vocal sequencing emerge while birds are also adaptively learning to change the pitch
442 of the optically targeted song syllable, we next examined if there was a relationship between adaptive (pitch
443 learning) and maladaptive (syllable repetitions) forms of motor plasticity. We found that optical inhibition
444 experiments drove changes in the pitch of song syllables comparable to those seen in optical excitation
445 experiments, yet these manipulations did not result in maladaptive vocal repetitions (Figures 5B, 5C and 6B).
446 This suggests that reinforcement-based learning of changes in pitch can occur independent of maladaptive
447 changes in syllable sequencing. In addition, we found that the recovery from optogenetic induced changes
448 in vocal repetitions occurred on timescales that differed from recovery in pitch learning (Figures 6D and 6E).
449 Recovery trajectories for these two types of learning varied from bird to bird, but their decoupling further
450 suggests that reinforcement-based changes in how syllables are sung (pitch learning) occurs independent of
451 maladaptive changes to higher-order representations that may help regulate vocal sequences.

452 Together, these findings indicate that increased phasic excitation of dopaminergic terminals in Area X cause
453 problems with initiating and terminating song, marked by birds perseverating on syllables at the beginning

454 and ending of song. That only phasic excitation is sufficient to drive these disruptions implicates the direct
455 pathway, and FoxP2 expressing neurons in this circuit, is involved in higher-order representations of song
456 sequences.

457



458

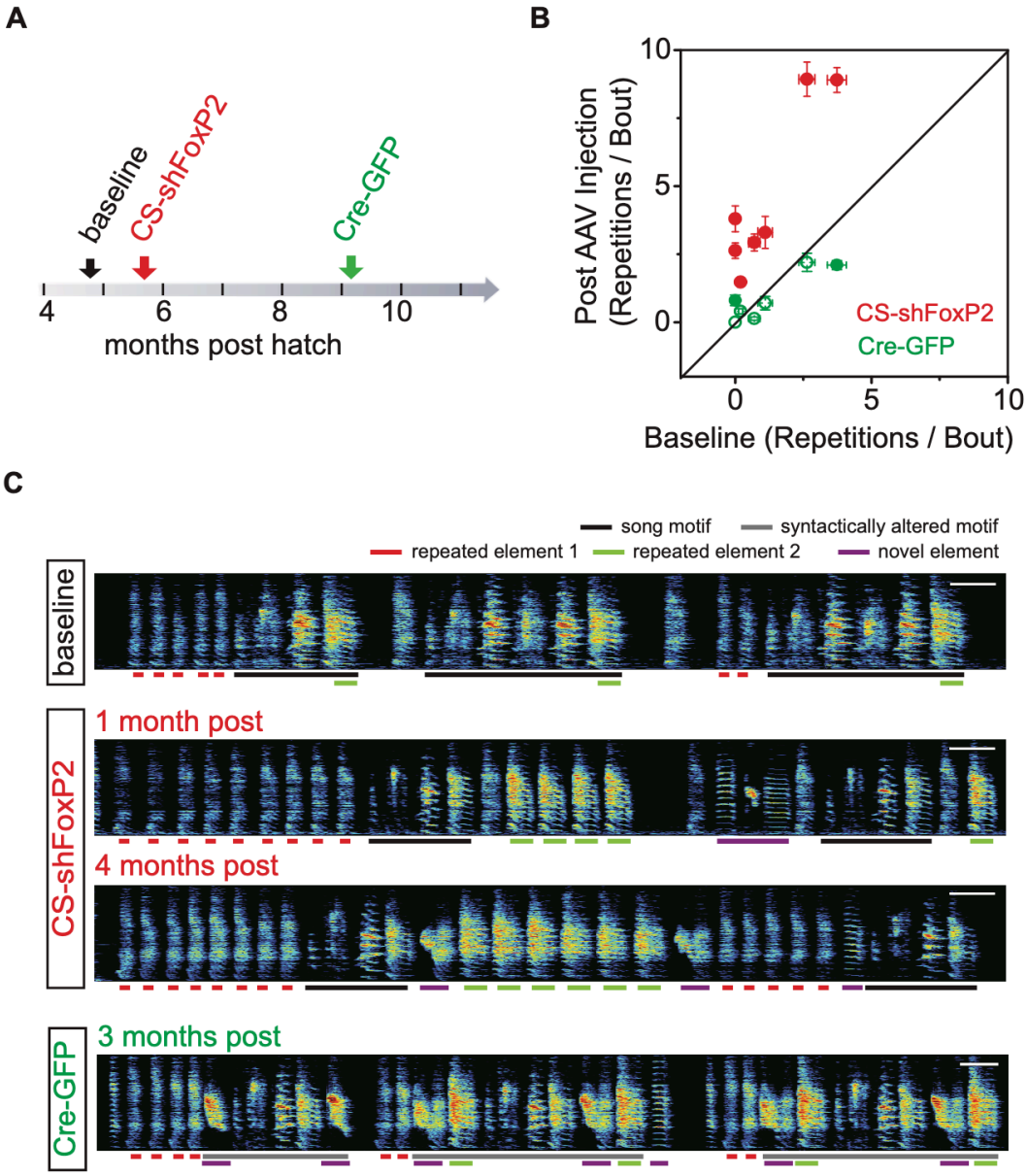
459 **Figure 6. Optical Stimulation Causes Vocal Repetitions Independent of Changes in Pitch.**
460 (A) Comparison of number of repetitions of vocal element per song bout (± SEM) for the baseline day versus last stimulation day
461 in 6 ChR2+ birds (filled, $p<0.0001$, Kruskal-Wallis test) or last recovery day (open, $p>0.35$, Kruskal-Wallis test, n=8 vocal
462 elements). The number of repetitions of vocal elements in the last stimulation day was significantly higher than the number of
463 repetitions in either the baseline day or the last recovery day ($p=0.0081$, Friedman test), but not between the baseline day and
464 the last recovery day ($p>0.99$, Friedman test).
465 (B) Comparison of number of repetitions of vocal element per song bout (± SEM) for the baseline day versus last inhibition day in
466 6 ArchT+ birds (filled, $p>0.08$; filled with black outline $p=0.028$, Kruskal-Wallis test, n=8 syllables).
467 (C) Same as (B), but for GFP+ birds (blue outline, $p>0.8$, Kruskal-Wallis test, n=6 syllables from 3 birds; orange outline, $p>0.15$,
468 Kruskal-Wallis test, n=7 syllables from 3 birds). Blue and orange outline indicate birds illuminated by LED with wavelength of
469 ~455 nm or 520 nm, respectively.
470 (D) The number of repetitions of a syllable from bird shown in Fig. 5D per song bout (top) and changes of mean pitch(bottom) of
471 the optically targeted syllable in Fig. 5D) during baseline (black circles, day -1 and 0), stimulation (blue circles, day 1-3), and
472 recovery days (open blue circles, day 4-9). The number of repetitions of the affected syllable (red line in Fig. 5D) per song bout
473 was significantly increased during day 1-6 relative to baseline ($p<0.01$, Kruskal-Wallis test), whereas changes in the mean pitch of
474 the target syllable during day 1- 9 were significantly higher than changes in the baseline day ($p<0.01$, Kruskal-Wallis test).

(E) same as (D), but for another bird. The number of repetitions of the affected syllable per song bout was significantly increased during day 2-8 relative to baseline ($p<0.01$, Kruskal-Wallis test), whereas changes in the mean pitch of the target syllable during day 1- 5 were significantly higher than changes relative to baseline ($p<0.01$, Kruskal-Wallis test).

Rescue of FoxP2 Knock-down Reverses Disruptions in Syllable Sequencing

Encouraged by the recovery of normal vocal sequencing following termination of our optical stimulation experiments, we tested if we could also rescue disruptions in the selection and sequencing of vocal-motor actions driven by long-term knockdown of FoxP2 in adult birds. We injected Area X with an AAV expressing Cre-GFP 2-6 months following injection of CS-shFoxP2 (AAVs expressing CS-shFoxP2 and Cre-GFP were injected at 177±24 dph and 277±23 dph, respectively, Figure 7A). Vocal disruptions in these birds were largely eliminated within 3 months following injection of Cre-GFP. Vocal repetitions were reduced to levels exhibited prior to knockdown of FoxP2 in all birds (Figures 7B, 7C and S5). During this slow recovery in vocal performance, some birds incorporated new vocal elements and changes in syntax that emerged during the knockdown period into their 'new' song. However, the large-scale disruptions marked by maladaptive vocal perseveration were rescued in all birds by reversal of the FoxP2 knockdown (Figures 7C and S5). These findings reveal remarkable adult plasticity in basal ganglia circuits controlling sequencing of vocal-motor actions and suggest that genetic intervention in adults may help correct speech disfluencies.

Figure 7 (next page). Reversal of FoxP2 KD in Adult Zebra Finches Rescues Disruptions in Song Syntax.
(A) Strategy to rescue the expression levels of FoxP2 in Area X of adult birds. AAV encoding Cre-GFP was bilaterally injected into Area X 2-6 months after the initial KD of FoxP2 with CS-shFoxP2 construct.
(B) Comparison of the number of repetitions of vocal element per song bout (± SEM) for the baseline day versus 2 months after injection of CS-shFoxP2 construct or 2 months after injection of Cre-GFP construct ($n=7$ vocal elements from 5 birds). Red filled circles represent vocal elements with significant differences between the baseline and post injection of CS-shFoxP2 ($p<0.0001$ and $p=0.012$, Kruskal-Wallis test). Green open circles represent vocal elements with nonsignificant differences ($p>0.2$, Kruskal- Wallis test), whereas green filled circles represent vocal elements with significant differences ($p=0.014$ and $p=0.027$, Kruskal-Wallis test) between the baseline and post injection of Cre-GFP. The number of repetitions of vocal elements after injection of Cre-GFP construct was significantly lower than the number of repetitions of the same elements after injection of CS-shFoxP2 construct ($p = 0.016$, $n = 7$, Wilcoxon matched-pairs signed-rank test).
(C) Spectrograms of song from one bird recorded on the baseline day, 1 month and 4 months after injection of CS-shFoxP2 construct, and 3 months after injection of Cre-GFP construct in Area X. Scale bar, 200 ms.



DISCUSSION

Precise sequencing of vocal motor actions is necessary for vocal communication. While recent studies have begun to clarify the role of Area X and reinforcement signals in learning how to properly produce individual syllables (Gadagkar et al., 2016; Hisey et al., 2018; Hoffmann et al., 2016; Xiao et al., 2018), the overall identity of cells and circuits that control selection and sequencing of syllables have remained unclear. Here we used RNA sequencing to catalog the molecular identity of Area X neurons, and then used genetic and optogenetic manipulations to test the behavioral function of this circuitry in regulating vocal sequences. We found numerous specializations in Area X, but also found that Area X striatal MSNs have a strong correspondence to ancestral direct and indirect pathways. In contrast, the intermingled pallidal neurons in Area X molecularly resemble arkypallidal cells of the GPe, a finding that may account for the known lack of projections between Area X and the STN.

We show that expression of FoxP2 is critical for the maintenance of adult vocalizations and that knockdown of its expression causes syllable repetitions and disruptions to song syntax. Restoring FoxP2 expression later in adulthood resulted in recovery of linear song syntax. This adult plasticity is particularly striking given that birds of this age are thought to have limited song plasticity and be less reliant on auditory feedback to maintain their songs(Lombardino and Nottebohm, 2000). We identify phasic dopamine as a selective mediator of maladaptive changes in song sequences and show that dopamine plays functions in reinforcement-based learning of song syllables and in the sequencing of those syllables. These findings demonstrate unexpected commonalities and specializations in basal ganglia circuits controlling learned vocalizations, and roles for FoxP2 and phasic dopamine in the maintenance of previously learned vocal motor sequences.

Many speech disorders arise from problems in translating volitional speech plans into accurate motor actions (Kang and Drayna, 2011; Konopka and Roberts, 2016; Krishnan et al., 2016; Newbury and Monaco, 2010) and have been linked to hyperdopaminergic signaling in the striatum (Alm, 2004; Anderson et al., 1999; Craig-McQuaide et al., 2014; Wu et al., 1997). Together, our findings indicate that convergent circuit mechanisms may be involved in translating volitional birdsong and speech plans into to fluent vocal-motor actions.

Evolution of Basal Ganglia Modules for Learned Vocalizations

We identified distinct populations of *Drd1*+/*FoxP2*+ MSNs and *Drd2*+/*FoxP2*- MSNs in Area X, consistent with striatal direct and indirect MSN populations seen across all vertebrates. Nearly 20% of the MSNs in Area X co-express *Drd1* and *Drd2*. Although this proportion is less than what has been previously reported, with most Area X MSNs thought to co-express both genes (Ding and Perkel, 2002; Kubikova et al., 2010), it is much higher than what has been reported in the mammalian striatum, where MSNs co-expressing *Drd1* and *Drd2* only represent about 1-4% of total MSNs (Anderson et al., 2020; Gokce et al., 2016; Saunders et al., 2018). Although the broad groupings of MSNs into direct and indirect pathways highlight commonalities in the gene markers between Area X and mammalian pathways (Anderson et al., 2020; Gokce et al., 2016; Ho et al., 2018), this is not the case for all genes. While *Drd2* and *FoxP2* are differentially expressed in MSN neuronal clusters, *Drd1* is not. Furthermore, a substantial portion (33%) of the MSNs in our data set do not express *any* dopamine receptor. Although this may reflect the inability to detect transcripts expressed at low levels, 33% is much higher than what has been reported in mammalian striatum (Anderson et al., 2020; Gokce et al., 2016; Saunders et al., 2018). Additionally, the nearly 30-fold difference in the numbers of MSNs (9,672 cells) compared to PNs (325 cells) indicates an enormous signaling convergence between the MSNs and PNs. Such convergence between striatal and pallidal cells, though anatomically segregated in mammals, is thought to be important for learning within the basal ganglia (Fee and Goldberg, 2011; Goldberg et al., 2013; Grillner and Robertson, 2016).

Previous studies have identified two cell types in Area X that share morphological and electrophysiological properties with cells seen in the mammalian GPi and GPe (Farries et al., 2005a; Goldberg et al., 2010).

561 Therefore, we expected to observe cell clusters corresponding to GPi and prototypical GPe neurons.
562 Instead, we found that pallidal neurons in Area X appear most similar to arkypallidal cells of the GPe.
563 Arkypallidal cells have only recently been described in the mammalian GPe. Although they comprise only
564 approximately one-quarter of GPe neurons (the other three-quarters being prototypical GPe cells), they
565 have been shown to provide a strong, inhibitory "stop" signal to the striatum (Abdi et al., 2015; Hegeman
566 et al., 2016; Mallet et al., 2012; Mallet et al., 2016). We propose that the thalamus-projecting neurons in
567 Area X are a novel specialization derived from an arkypallidal cell-type progenitor. The presence of a novel
568 arkypallidal-like cell-type in the striatal region Area X may indicate a novel evolutionary modification of
569 striato-pallidal basal ganglia circuits to facilitate the learning and production of a highly specialized behavior
570 (song).
571
572 Although the basal ganglia are strongly conserved across vertebrates, some differences in cell-types and
573 connectivity do exist between taxa. Area X is a specialization of the dorsal medial striatum, with both
574 regions containing many of the same cell-types (Person et al., 2008). At least one pallidal-like cell-type has
575 been reported in the medial striatum of songbirds outside of Area X (Person et al., 2008), and a similar but
576 still distinct cell-type has been described in the medial striatum of a non-songbird (a chicken) (Farries et al.,
577 2005b). The relationship between these various cell types is not yet clear. Arkypallidal cells are
578 developmentally derived from the progenitor of the medial striatum, the lateral ganglionic eminence
579 (Hegeman et al., 2016; Medina et al., 2014; Nóbrega-Pereira et al., 2010). Our interpretation suggests that,
580 in birds, the medial striatum retains a primordial arkypallidal-like cell-type, which in songbirds has further
581 specialized into a specific cell-type for vocal circuits. The rapid transitions between syllables during singing
582 may have necessitated fast pathways that function to stop the ongoing syllable or its repetition and allow
583 transitioning to the next syllable.
584

585 **Role of FoxP2 in Adult Vocal Motor Control**

586 Heterozygous mutations of FOXP2 cause Childhood Apraxia of Speech, also referred to as Developmental
587 Verbal Dyspraxia. This speech impairment is thought to result in part from disruptions in developmental
588 plasticity of basal ganglia circuits (Ullman et al., 2020). The role of FOXP2 in the maintenance of adult speech
589 is not known. Altering the expression of FoxP2 in Area X impairs song imitation in juvenile birds, but it had
590 previously not been shown to be necessary for maintenance of adult song. Using the same hairpin sequence
591 used in previous studies(Haesler et al., 2007; Murugan et al., 2013), but expressing it in Area X using a novel
592 AAV CS-shFoxP2 construct, we show that FoxP2 expression is necessary for maintenance of adult song
593 sequences and syntax. Knocking down FoxP2 expression in Area X of adult zebra finches drove a significant
594 increase in the repetition of song syllables, elimination of certain syllables from the song, and the
595 improvisation of new syllables. However, we did not observe degradation of individual song syllables, as
596 occurs following deafening. This indicates that the ability to control fine-scale features of song was largely
597 undisturbed, while global control of selection and sequencing of syllables was impaired.

598 We and others find that knockdown of FoxP2 results in a decreased expression of *Drd1*/5, and it has
599 previously been shown that knockdown speeds dopamine-sensitive signal transmission through the basal

ganglia (Murugan et al., 2013). One possibility is that imbalances between the direct and indirect pathways, and faster signal transmission through the direct pathway, disrupt the timing of syllable-level cancellation signals that may arise from pallidal circuits to facilitate precise transitioning from one syllable to the next.

When we rescued FoxP2 knockdown in adult birds, we found that they were able to recover normal song syntax within ~3 months-time. This indicates that the progressive and maladaptive changes in behavior driven by disruptions of FoxP2 can be overcome with restoration of gene expression. Despite its name, Childhood Apraxia of Speech is a lifelong condition. Finding that birds can recover normal song syntax well into adulthood, a timepoint when song behavior is thought to be mostly rote (Lombardino and Nottebohm, 2000), suggests the possibility that genetic therapies for speech disorders could have relevance even beyond early developmental windows when speech is first learned.

Dual Role of Area X in Reinforcement Learning and Controlling Vocal-motor Sequences

Our findings indicate that Area X plays a dual function. It is involved in learning how individual song syllables should be sung and in controlling larger scale selection and sequencing of these syllables. Positive and negative dopaminergic reinforcement signals guide how individual song syllables are sung on future performances, while disruptions to the contributions of the direct and indirect pathways may regulate syllable selection and sequencing. As mentioned above, knockdown of FoxP2 leads to a decrease in *Drd1*/*Drd2* ratios, which may drive an imbalance in direct and indirect pathways. Moreover, continued phasic excitation of dopaminergic inputs, which is thought to preferentially influence activity in the direct pathway, also drives disruptions in song sequencing. Our manipulations of the direct pathway most consistently resulted in birds having prominent sequence disruptions/repetitions at the beginning and end of their song.

Similarly, expression of the mutant gene fragment that causes Huntington's disease in Area X also causes disruptions in song syllable selection and repetition (Tanaka et al., 2016). These disruptions, however, tend to be more restricted to changes in core aspects of the song and do not accumulate at the initiation and termination of song. Indirect pathway neurons are particularly vulnerable at early stages of Huntington's disease (Albin et al., 1992; Reiner et al., 1988; Richfield et al., 1995), which together with our findings, raises speculation that disruptions in the direct pathway could more readily cause vocal repetitions at initiation and termination of vocal-motor sequences, whereas disruptions in the indirect pathway could tend to disrupt sequences in the middle of song. Hierarchical representations of song sequences may therefore rely critically on coordinated activity of the direct and indirect pathways in Area X and the precise timing signals that facilitate transitions between individual syllables. The molecular cataloging of Area X cell types, and the tools for reversible genetic manipulations described here, provide the means to start testing these and related hypotheses about the selection and sequencing of vocal-motor actions.

Acknowledgments: The authors thank members of the Roberts and Konopka laboratories for discussion and comments on the manuscript, Jennifer Holdway and Matthew Harper for laboratory support, Garav Chattree for helping with optogenetic experiments, Maaya Ikeda and Chung Yan Cheung for advice on reverse microdialysis experiments, Andrea Guerrero for recording of directed singing behavior, Ashley Anderson for cloning of FoxP2 V5, and Erich Jarvis for the original clone of zebra finch FoxP2. **Funding:** This research was supported by grants from the US National Institutes of Health R21DC016340 to TFR and GK, R01NS102488 to TFR and R01DC014702 to GK. DPM was supported by F32NS112557. **Author contributions:** LX, DPM. and TFR designed the experiments and wrote the manuscript, LX collected and analyzed the optogenetic, pharmacological, and gene knockdown experiments, and help collect the snRNA-seq data, DPM analyzed the snRNA-seq data, MC^2 analyzed and imaged the anatomical data and helped interpret the directed singing data, MC^1 collected the snRNAseq data, AK developed bioinformatic pipeline for snRNA-seq data analysis and helped analyze the data, GK supervised the snRNA-seq data collection and analysis and helped design the reversible gene knockdown experiments, TFR supervised all experiments. All authors read and commented on the manuscript. **Competing interests:** Authors declare no competing interests. **Data and materials availability:** All data is available in the main text or the supplementary materials.

Supplemental Information:

Materials and Methods

Supplemental Figures 1 to 7

REFERENCES

Abdi, A., Mallet, N., Mohamed, F.Y., Sharott, A., Dodson, P.D., Nakamura, K.C., Suri, S., Avery, S.V., Larvin, J.T., Garas, F.N.*, et al.* (2015). Prototypic and arkypallidal neurons in the dopamine-intact external globus pallidus. J Neurosci *35*, 6667.

Albin, R.L., Reiner, A., Anderson, K.D., Dure, L.S.t., Handelin, B., Balfour, R., Whetsell, W.O., Jr., Penney, J.B., and Young, A.B. (1992). Preferential loss of striato-external pallidal projection neurons in presymptomatic Huntington's disease. Ann Neurol *31*, 425-430.

Albin, R.L., Young, A.B., and Penney, J.B. (1995). The functional anatomy of disorders of the basal ganglia. Trends Neurosci *18*, 63-64.

Alm, P.A. (2004). Stuttering and the basal ganglia circuits: a critical review of possible relations. Journal of communication disorders *37*, 325-369.

Anderson, A.G., Kulkarni, A., Harper, M., and Konopka, G. (2020). Single-cell analysis of Foxp1-driven mechanisms essential for striatal development. Cell Reports *30*, 3051-3066.e3057.

Anderson, J.M., Hughes, J.D., Rothi, L.J., Crucian, G.P., and Heilman, K.M. (1999). Developmental stuttering and Parkinson's disease: the effects of levodopa treatment. J Neurol Neurosurg Psychiatry *66*, 776-778.

Bottjer, S.W., Miesner, E.A., and Arnold, A.P. (1984). Forebrain lesions disrupt development but not maintenance of song in passerine birds. Science *224*, 901-903.

Budzillo, A., Duffy, A., Miller, K.E., Fairhall, A.L., and Perkel, D.J. (2017). Dopaminergic modulation of basal ganglia output through coupled excitation–inhibition. PNAS *114*, 5713.

675 Carrillo, G.D., and Doupe, A.J. (2004). Is the songbird Area X striatal, pallidal, or both? An anatomical study.
676 J Comp Neurol *473*, 415-437.
677 Chen, J.Y., Wang, E.A., Cepeda, C., and Levine, M.S. (2013). Dopamine imbalance in Huntington's disease: a
678 mechanism for the lack of behavioral flexibility. Frontiers in neuroscience *7*, 114.
679 Chen, Y.-J.J., Friedman, B.A., Ha, C., Durinck, S., Liu, J., Rubenstein, J.L., Seshagiri, S., and Modrusan, Z.
680 (2017). Single-cell RNA sequencing identifies distinct mouse medial ganglionic eminence cell types. Sci Rep
681 *7*, 45656.
682 Coddington, L.T., and Dudman, J.T. (2019). Learning from Action: Reconsidering Movement Signaling in
683 Midbrain Dopamine Neuron Activity. Neuron *104*, 63-77.
684 Craig-McQuaide, A., Akram, H., Zrinzo, L., and Tripoliti, E. (2014). A review of brain circuitries involved in
685 stuttering. Frontiers in human neuroscience *8*, 884.
686 Day, N.F., Hobbs, T.G., Heston, J.B., and White, S.A. (2019). Beyond Critical Period Learning: Striatal FoxP2
687 Affects the Active Maintenance of Learned Vocalizations in Adulthood. eNeuro *6*.
688 DeLong, M.R. (1990). Primate models of movement disorders of basal ganglia origin. Trends Neurosci *13*,
689 281-285.
690 Ding, L., and Perkel, D.J. (2002). Dopamine modulates excitability of spiny neurons in the avian basal
691 ganglia. J Neurosci *22*, 5210.
692 Ding, L., Perkel, D.J., and Farries, M.A. (2003). Presynaptic depression of glutamatergic synaptic
693 transmission by D1-like dopamine receptor activation in the avian basal ganglia. J Neurosci *23*, 6086-6095.
694 Dreyer, J.K., Herrik, K.F., Berg, R.W., and Hounsgaard, J.D. (2010). Influence of phasic and tonic dopamine
695 release on receptor activation. J Neurosci *30*, 14273-14283.
696 Enard, W., Gehre, S., Hammerschmidt, K., Holter, S.M., Blass, T., Somel, M., Bruckner, M.K., Schreiweis, C.,
697 Winter, C., Sohr, R.*, et al.* (2009). A humanized version of Foxp2 affects cortico-basal ganglia circuits in
698 mice. Cell *137*, 961-971.
699 Farries, M.A., Ding, L., and Perkel, D.J. (2005a). Evidence for “direct” and “indirect” pathways through the
700 song system basal ganglia. J Comp Neurol *484*, 93-104.
701 Farries, M.A., Meitzen, J., and Perkel, D.J. (2005b). Electrophysiological properties of neurons in the basal
702 ganglia of the domestic chick: Conservation and divergence in the evolution of the avian basal ganglia. J
703 Neurophysiol *94*, 454-467.
704 Fee, M.S., and Goldberg, J.H. (2011). A hypothesis for basal ganglia-dependent reinforcement learning in
705 the songbird. Neuroscience *198*, 152-170.
706 Fisher, S.E., Lai, C.S.L., Hurst, J.A., Vargha-Khadem, F., and Monaco, A.P. (2001). A novel forkhead-domain
707 gene is mutated in a severe speech and language disorder. Am J Hum Genet *69*, 176-176.
708 Gadagkar, V., Puzerey, P.A., Chen, R., Baird-Daniel, E., Farhang, A.R., and Goldberg, J.H. (2016). Dopamine
709 neurons encode performance error in singing birds. Science *354*, 1278-1282.
710 Gale, S.D., and Perkel, D.J. (2010). Anatomy of a songbird basal ganglia circuit essential for vocal learning
711 and plasticity. J Chem Neuroanat *39*, 124-131.
712 Girasole, A.E., Lum, M.Y., Nathaniel, D., Bair-Marshall, C.J., Guenthner, C.J., Luo, L., Kreitzer, A.C., and
713 Nelson, A.B. (2018). A Subpopulation of Striatal Neurons Mediates Levodopa-Induced Dyskinesia. Neuron
714 *97*, 787-795 e786.

715 Gokce, O., Stanley, Geoffrey M., Treutlein, B., Neff, Norma F., Camp, J.G., Malenka, Robert C., Rothwell,
716 Patrick E., Fuccillo, Marc V., Südhof, Thomas C., and Quake, Stephen R. (2016). Cellular taxonomy of the
717 mouse striatum as revealed by single-cell RNA-seq. Cell Reports *16*, 1126-1137.
718 Goldberg, J.H., Adler, A., Bergman, H., and Fee, M.S. (2010). Singing-related neural activity distinguishes
719 two putative pallidal cell types in the songbird basal ganglia: Comparison to the primate internal and
720 external pallidal segments. J Neurosci *30*, 7088.
721 Goldberg, J.H., Farries, M.A., and Fee, M.S. (2013). Basal ganglia output to the thalamus: Still a paradox.
722 Trends Neurosci *36*, 695-705.
723 Grillner, S., and Robertson, B. (2016). The Basal Ganglia Over 500 Million Years. Curr Biol *26*, R1088-R1100.
724 Grillner, S., Robertson, B., and Stephenson-Jones, M. (2013). The evolutionary origin of the vertebrate basal
725 ganglia and its role in action selection. J Physiol-London *591*, 5425-5431.
726 Haesler, S., Rochefort, C., Georgi, B., Licznerski, P., Osten, P., and Scharff, C. (2007). Incomplete and
727 inaccurate vocal imitation after knockdown of FoxP2 in songbird basal ganglia nucleus Area X. PLoS Biol *5*,
728 e321.
729 Hegeman, D.J., Hong, E.S., Hernández, V.M., and Chan, C.S. (2016). The external globus pallidus: Progress
730 and perspectives. Eur J Neurosci *43*, 1239-1265.
731 Hikosaka, O., Nakahara, H., Rand, M.K., Sakai, K., Lu, X., Nakamura, K., Miyachi, S., and Doya, K. (1999).
732 Parallel neural networks for learning sequential procedures. Trends Neurosci *22*, 464-471.
733 Hisey, E., Kearney, M.G., and Mooney, R. (2018). A common neural circuit mechanism for internally guided
734 and externally reinforced forms of motor learning. Nat Neurosci *21*, 589-+.
735 Ho, H., Both, M.D., Siniard, A., Sharma, S., Notwell, J.H., Wallace, M., Leone, D.P., Nguyen, A., Zhao, E., Lee,
736 H.*, et al.* (2018). A guide to single-cell transcriptomics in adult rodent brain: The medium spiny neuron
737 transcriptome revisited. Frontiers in Cellular Neuroscience *12*.
738 Hoffmann, L.A., Saravanan, V., Wood, A.N., He, L., and Sober, S.J. (2016). Dopaminergic Contributions to
739 Vocal Learning. J Neurosci *36*, 2176-2189.
740 Jin, X., and Costa, R.M. (2015). Shaping action sequences in basal ganglia circuits. Curr Opin Neurobiol *33*,
741 188-196.
742 Kang, C., and Drayna, D. (2011). Genetics of speech and language disorders. Annu Rev Genomics Hum
743 Genet *12*, 145-164.
744 Klaus, A., da Silva, J.A., and Costa, R.M. (2019). What, If, and When to Move: Basal Ganglia Circuits and Self-
745 Paced Action Initiation. Annual Review of Neuroscience, Vol 42 *42*, 459-483.
746 Konopka, G., and Roberts, T.F. (2016). Insights into the neural and genetic basis of vocal communication.
747 Cell *164*, 1269-1276.
748 Krishnan, S., Watkins, K.E., and Bishop, D.V. (2016). Neurobiological Basis of Language Learning Difficulties.
749 Trends Cogn Sci *20*, 701-714.
750 Kubikova, L., Wada, K., and Jarvis, E.D. (2010). Dopamine receptors in a songbird brain. J Comp Neurol *518*,
751 741-769.
752 Lai, C.S.L., Fisher, S.E., Hurst, J.A., Vargha-Khadem, F., and Monaco, A.P. (2001). A forkhead-domain gene is
753 mutated in a severe speech and language disorder. Nature *413*, 519-523.
754 Lombardino, A.J., and Nottebohm, F. (2000). Age at deafening affects the stability of learned song in adult
755 male zebra finches. J Neurosci *20*, 5054-5064.

Mallet, N., Micklem, B.R., Henny, P., Brown, M.T., Williams, C., Bolam, J.P., Nakamura, K.C., and Magill, P.J. (2012). Dichotomous organization of the external globus pallidus. Neuron *74*, 1075-1086.

Mallet, N., Schmidt, R., Leventhal, D., Chen, F., Amer, N., Boraud, T., and Berke, Joshua D. (2016). Arkypallidal cells send a stop signal to striatum. Neuron *89*, 308-316.

Medina, L., Abellán, A., Vicario, A., and Desfilis, E. (2014). Evolutionary and developmental contributions for understanding the organization of the basal ganglia. Brain Behav Evol *83*, 112-125.

Mendoza, E., Tokarev, K., During, D.N., Retamosa, E.C., Weiss, M., Arpenik, N., and Scharff, C. (2015). Differential coexpression of FoxP1, FoxP2, and FoxP4 in the Zebra Finch (Taeniopygia guttata) song system. J Comp Neurol *523*, 1318-1340.

Murugan, M., Harward, S., Scharff, C., and Mooney, R. (2013). Diminished FoxP2 levels affect dopaminergic modulation of corticostriatal signaling important to song variability. Neuron *80*, 1464-1476.

Newbury, D.F., and Monaco, A.P. (2010). Genetic advances in the study of speech and language disorders. Neuron *68*, 309-320.

Nóbrega-Pereira, S., Gelman, D., Bartolini, G., Pla, R., Pierani, A., and Marín, O. (2010). Origin and molecular specification of globus pallidus neurons. J Neurosci *30*, 2824.

Nordeen, K.W., and Nordeen, E.J. (1993). Long-Term Maintenance of Song in Adult Zebra Finches Is Not Affected by Lesions of a Forebrain Region Involved in Song Learning. Behavioral and Neural Biology *59*, 79-82.

Norton, P., Barschke, P., Scharff, C., and Mendoza, E. (2019). Differential Song Deficits after Lentivirus-Mediated Knockdown of FoxP1, FoxP2, or FoxP4 in Area X of Juvenile Zebra Finches. J Neurosci *39*, 9782-9796.

Penhune, V.B., and Steele, C.J. (2012). Parallel contributions of cerebellar, striatal and M1 mechanisms to motor sequence learning. Behav Brain Res *226*, 579-591.

Person, A.L., Gale, S.D., Farries, M.A., and Perkel, D.J. (2008). Organization of the songbird basal ganglia, including Area X. J Comp Neurol *508*, 840-866.

Reiner, A. (1987). A LANT6-like substance that is distinct from neuromedin N is present in pallidal and striatal neurons in monkeys. Brain Res *422*, 186-191.

Reiner, A. (2016). The conservative evolution of the vertebrate basal ganglia. In Handbook of Behavioral Neuroscience, H. Steiner, and K.Y. Tseng, eds. (Elsevier), pp. 63-97.

Reiner, A., Albin, R.L., Anderson, K.D., D'Amato, C.J., Penney, J.B., and Young, A.B. (1988). Differential loss of striatal projection neurons in Huntington disease. Proc Natl Acad Sci U S A *85*, 5733-5737.

Reiner, A., and Carraway, R.E. (1987). Immunohistochemical and biochemical studies on Lys8-Asn9-neurotensin8–13(LANT6)-related peptides in the basal ganglia of pigeons, turtles, and hamsters. J Comp Neurol *257*, 453-476.

Reiner, A., Laverghetta, A.V., Meade, C.A., Cuthbertson, S.L., and Bottjer, S.W. (2004). An immunohistochemical and pathway tracing study of the striatopallidal organization of Area X in the male zebra finch. J Comp Neurol *469*, 239-261.

Richfield, E.K., Maguire-Zeiss, K.A., Vonkeman, H.E., and Voorn, P. (1995). Preferential loss of preproenkephalin versus preprotachykinin neurons from the striatum of Huntington's disease patients. Ann Neurol *38*, 852-861.

796 Richfield, E.K., Penney, J.B., and Young, A.B. (1989). Anatomical and affinity state comparisons between
797 dopamine D1 and D2 receptors in the rat central nervous system. Neuroscience *30*, 767-777.
798 Roberts, T.F., Tschida, K.A., Klein, M.E., and Mooney, R. (2010). Rapid spine stabilization and synaptic
799 enhancement at the onset of behavioural learning. Nature *463*, 948-952.
800 Saunders, A., Johnson, C.A., and Sabatini, B.L. (2012). Novel recombinant adeno-associated viruses for Cre
801 activated and inactivated transgene expression in neurons. Frontiers in neural circuits *6*, 47.
802 Saunders, A., Macosko, E.Z., Wysoker, A., Goldman, M., Krienen, F.M., de Rivera, H., Bien, E., Baum, M.,
803 Bortolin, L., Wang, S.*, et al.* (2018). Molecular diversity and specializations among the cells of the adult
804 mouse brain. Cell *174*, 1015-1030.e1016.
805 Scharff, C., and Nottebohm, F. (1991). A Comparative-Study of the Behavioral Deficits Following Lesions of
806 Various Parts of the Zebra Finch Song System - Implications for Vocal Learning. J Neurosci *11*, 2896-2913.
807 Simonyan, K. (2019). Recent advances in understanding the role of the basal ganglia. F1000Res *8*, F1000
808 Faculty Rev-1122.
809 Sohrabji, F., Nordeen, E.J., and Nordeen, K.W. (1990). Selective impairment of song learning following
810 lesions of a forebrain nucleus in the juvenile zebra finch. Behav Neural Biol *53*, 51-63.
811 Stephenson-Jones, M., Samuelsson, E., Ericsson, J., Robertson, B., and Grillner, S. (2011). Evolutionary
812 Conservation of the Basal Ganglia as a Common Vertebrate Mechanism for Action Selection. Current
813 Biology *21*, 1081-1091.
814 Tanaka, M., Singh Alvarado, J., Murugan, M., and Mooney, R. (2016). Focal expression of mutant huntingtin
815 in the songbird basal ganglia disrupts cortico-basal ganglia networks and vocal sequences. Proc Natl Acad
816 Sci U S A *113*, E1720-1727.
817 Tepper, J.M., and Bolam, J.P. (2004). Functional diversity and specificity of neostriatal interneurons. Curr
818 Opin Neurobiol *14*, 685-692.
819 Tepper, J.M., Tecuapetla, F., Koós, T., and Ibáñez-Sandoval, O. (2010). Heterogeneity and diversity of
820 striatal GABAergic interneurons. Front Neuroanat *4*, 150-150.
821 Ullman, M.T., Earle, F.S., Walenski, M., and Janacsek, K. (2020). The Neurocognition of Developmental
822 Disorders of Language. Annu Rev Psychol *71*, 389-417.
823 Vernes, S.C., Oliver, P.L., Spiteri, E., Lockstone, H.E., Puliyadi, R., Taylor, J.M., Ho, J., Mombereau, C.,
824 Brewer, A., Lowy, E.*, et al.* (2011). Foxp2 regulates gene networks implicated in neurite outgrowth in the
825 developing brain. PLoS Genet *7*, e1002145.
826 Wallace, M.L., Saunders, A., Huang, K.W., Philson, A.C., Goldman, M., Macosko, E.Z., McCarroll, S.A., and
827 Sabatini, B.L. (2017). Genetically distinct parallel pathways in the entopeduncular nucleus for limbic and
828 sensorimotor output of the basal ganglia. Neuron *94*, 138-152.e135.
829 Wu, J.C., Maguire, G., Riley, G., Lee, A., Keator, D., Tang, C., Fallon, J., and Najafi, A. (1997). Increased
830 dopamine activity associated with stuttering. Neuroreport *8*, 767-770.
831 Wymbs, N.F., Bassett, D.S., Mucha, P.J., Porter, M.A., and Grafton, S.T. (2012). Differential recruitment of
832 the sensorimotor putamen and frontoparietal cortex during motor chunking in humans. Neuron *74*, 936-
833 946.
834 Xiao, L., Chattree, G., Oscos, F.G., Cao, M., Wanat, M.J., and Roberts, T.F. (2018). A Basal Ganglia Circuit
835 Sufficient to Guide Birdsong Learning. Neuron *98*, 208-+.

Yokoi, A., and Diedrichsen, J. (2019). Neural Organization of Hierarchical Motor Sequence Representations in the Human Neocortex. Neuron *103*, 1178-1190 e1177.

Yu, X., Ye, Z.W., Houston, C.M., Zecharia, A.Y., Ma, Y., Zhang, Z., Uygun, D.S., Parker, S., Vyssotski, A.L., Yustos, R.*, et al.* (2015). Wakefulness Is Governed by GABA and Histamine Cotransmission. Neuron *87*, 164-178.

Supplementary Materials: Functional Identification of Specialized Basal Ganglia Circuits that Regulate Vocal Motor Sequences

Materials and Methods

Figures S1 to S7

Supplemental References

Materials and Methods

Animals

All experiments were performed on adult male zebra finches (*Taeniopygia guttata*) raised in a breeding facility at UT Southwestern and housed with their parents until at least 50 days of age. During experiments, birds were housed individually in sound-attenuating recording chambers (Med associates) on a 12/12 h day/night cycle and were given ad libitum access to food and water. All procedures were performed in accordance with established protocols approved by the UT Southwestern Medical Center Animal Care and Use Committee.

Plasmid Construction and Viral Vectors

The backbone of CS constructs was based on pAAV-EF1α-DO-mCherry (Addgene, #37119)(Saunders et al., 2012), and the fluorescent protein cDNA for tagBFP was cloned from pdCas9::BFP-humanized (Addgene, #44247). A zebra finch FoxP2 cDNA clone provided by Erich Jarvis was subcloned into pLenti6.4 using a gateway reaction, adding a V5 tag. Two hairpins (shFoxP2a, target sequence AACAGGAAGCCCAACGTTAG T(Haesler et al., 2007), and shFoxP2i, target sequence ACTCATCATTCCATAGTGAAT) were inserted downstream of the U6 promoter at the base of the Mir-30 stem-loop. The scrambled hairpin (shScramble, sequence CCACTGTACTATCTATAACAT) was designed as a control. Hairpins were then assembled into the pTripZ vector (Thermo Scientific, MA, USA) by directional ligation into the XhoI-EcoRI cloning sites. We then replaced the EF1α promoter with a CAG promoter and assembled hairpins together with pTripZ context sequence (between the BspD1-MluI cloning sites) in the forward orientation and tagBFP transgene in the reversed orientation downstream of the mCherry transgene. All N Terminal sites included a Kozak sequence (GCCACC) directly preceding the start codon. Sequence confirmation was done by the McDermott Center Sequencing Core at UT Southwestern Medical Center. The recombinant AAV vectors were amplified by recombination deficient bacteria, One Shot Stbl3(C737303, Invitrogen, CA, USA), serotyped with AAV1 coat proteins and produced by the University of North Carolina vector core facility (Chapel Hill, NC, USA) with titer exceeding 10^{12} vg/ml, the Duke viral vector core facility (Durham, NC, USA), IDDRC Neuroconnectivity Core in Baylor College of Medicine (Huston, TX, USA) or in the Roberts lab with titer exceeding 10^{11} vg/ml. All viral vectors were aliquoted and stored at –80 °C until use.

Stereotaxic Surgery

All surgical procedures were performed under aseptic conditions. Birds were anesthetized using isoflurane inhalation (1.5-2%) and placed in a stereotaxic apparatus. Viral injections and cannula or microdialysis probe implantation were performed using previously described procedures(Xiao et al., 2018) at the following approximate stereotaxic coordinates relative to interaural zero and the brain surface were (rostral, lateral, depth, in mm): Ov (2.8, 1.0, 4.75), the center of Ov was located and mapped based on its robust white noise responses; VTA relative to the center of Ov (+0.3, -0.2, +1.8); Area X (5.1, 1.6, 3.3) with 43-degree head angle or (5.7, 1.6, 3) with 20-degree head angle, the boundary of Area X was verified using extracellular electrophysiological recordings. 0.7-2 µl AAVs were injected according to the titer of constructs and allowed 3-8 weeks for expression before birds were subjected for behavioral tests, immunohistochemistry and/or sequencing experiments.

Behavioral assays.

Song Recording

Acoustic signals were recorded continuously by a microphone immediately adjacent to the bird's cage using Sound Analysis Pro2011(Tchernichovski et al., 2000) and bandpass filtered between 0.3 and 10 kHz. All songs presented were recorded when the male was isolated in a sound-attenuating chamber.

Optogenetic Manipulation of VTA Axon Terminals in Behaving Birds

All procedures were reported previously (Xiao et al., 2018). Briefly, male birds were randomly assigned bilateral injection of either AAV-ChR2, -ArchT or GFP constructs in VTA at ~70 days post hatch(dph). Birds were implanted with fiber optics after they were at least ~100 dph. Birds were given at least 1 week to recover from cannula implantation and to habituate to sing with attached optical fibers. Custom LabView software (National Instruments) was used for online detection of predetermined target syllables and implementation of closed-loop optogenetic manipulation(Ali et al., 2013). 100-ms light pulses were delivered over a subset of variants of the target syllables in real time (system delay less than 25ms) for 3-12 consecutive days, as described previously(Xiao et al., 2018). Investigators were not blinded to allocation of optogenetic experiments. 3-5 mw of ~455nm and 1.5-4 mw of ~520nm LED output was delivered from the tip of the probe (200 or 250um, NA=0.66, Prizmatix, Israel) to ChR2+ and ArchT+ birds, respectively, while either ~455nm or ~520nm light pulses were delivered to GFP+ birds. ChR2+(n=6) and ArchT+ (n=6) birds with significant changes in the pitch of target syllables ($|d'| > 0.75$ significance threshold) were included for subsequent behavioral analyses (Figure 5C).

Pharmacological Manipulation of DA Circuit in Behaving Birds

We used two microdialysis systems to chronically infuse dopamine hydrochloride (DA), SKF 38393 hydrobromide (SKF) or (-)-Quinpirole hydrochloride (Qui) (#3548, #0922 and #106, Tocris, MN, USA) into Area X to simulate tonically elevated DA levels. Two male adult birds were implanted bilaterally with probes constructed in house from plastic tubing (427405, BD Intramedic, PA, USA; 27223 and 30006, MicroLumen, FL, USA) which served as a drug reservoir, fitted at the end with a 0.7mm-long semipermeable membrane (132294, Spectra/Por, MA, USA) allowing drug to slowly diffuse into the brain throughout the day(Hamaguchi and Mooney, 2012; Roberts et al., 2012). Freshly made DA, SKF or Qui(400mM) were used to fill and refilled microdialysis probes every morning for 4 consecutive days following 3 days of dialysis with PBS. Two other birds were implanted with guide cannula (8010684, CMA, MA, USA) bilaterally over Area X and microdialysis probes (1mm membrane length, 6kDa cutoff, P000082, CMA, MA, USA) were not inserted until birds recovered from surgery and were singing (2-3 days) as described previously(Miller et al., 2017; Tian and Brainard, 2017). Fresh made DA, SKF or Qui(100-400mM) or PBS (for baseline) were continuously delivered to Area X for 3 – 8 consecutive days at a rate of 0.2 µl/min via a fluid commutator connected to a syringe pump outside the bird's isolation chamber. In all cases birds could comfortably move and sing during infusion.

Behavioral analysis

Song structure. Zebra finch song can be classified into three levels of organization: syllables, which are individual song elements separated by short silent gaps >5 ms in duration; motifs, which are stereotyped sequences of syllables (outlined by black lines throughout the manuscript); and song bouts, which are defined as periods of singing comprised of introductory elements (eg. syllables indicated by red lines in Figure 4E), followed by one or more repeats of the song motif with inter-motif intervals >500 ms(Kao and Brainard, 2006; Sakata et al., 2008).

Quantification of vocal repetition. The vocal element being repeated consisted of either a single syllable (e.g. introductory element repeated in Figure 4E or ending song syllable repeated in Figure 4F) or multiple syllables (e.g. song syllables repeated in Figure S7C). The number of repetitions of individual vocal element per song bout(n) is defined as the total number of consecutively repeated vocal elements, not including the first rendition each time the element is sequentially produced within the song bout. To calculate the number of motifs per bout(m), we counted all motifs in which at least the first half of the motif was produced within a song bout. The number of repetitions of individual vocal elements per motif is defined as n/m. d'scores were computed to express the changes in the mean number of repetitions of individual vocal element per song bout(n) relative to the last baseline day(Xiao et al., 2018):

$$d_i' = \frac{\sqrt{2}(n_i - n_b)}{\sqrt{\sigma_i^2 + \sigma_b^2}}$$

n_i is the mean number of repetitions of individual vocal element per song bout on day *i* and σ_i^2 is the variance on day *i*. Subscript b refers to last baseline day. In the case of equal variances ($\sigma_i^2 = \sigma_b^2$), d_i' reports the changes in average repeats per bout between training day *i* and the baseline day in the convenient unit of SDs. Zebra finch song is mostly linear, exhibiting few repetitions. However, birds do tend to repeat introductory elements and ending elements a small number of times. We focused analysis on vocal elements or song syllables that were observed to repeat two or more times on at least one occasion during optical manipulations or following knockdown of FoxP2 in Area X. Once identified, we retrospectively tracked these syllables back to baseline periods or forward in time to recovery periods in order to examine if repetitions changed during our manipulations. Syllables which were not repeated during optical manipulations or following knockdown of FoxP2 were not further considered.

Immunohistochemistry and immunoblotting

Birds were anesthetized with Euthasol (Virbac, TX, USA) and transcardially perfused with ice cold phosphate buffered saline (PBS), followed by 4% paraformaldehyde (PFA) in PBS. The brains were post-fixed in 4% PFA for 2hrs at 4 °C, then transferred to PBS containing 0.05% sodium azide. The brains were sectioned at 50 µm, using a Leica VT1000S vibratome. Immunohistochemistry (IHC) was performed as described previously(Xiao et al., 2018). Sections were first washed in PBS, then blocked in 10% normal donkey serum (NDS) in PBST (0.3% Triton X-100 in PBS) for 1hr at room temperature (RT). Sections were incubated with primary antibodies in blocking solution (PBST with 2% NDS and 0.05% sodium azide) for 24-48 hrs at 4 °C, then washed in PBS before a secondary antibody incubation (703-545-155, 711-475-152, 705-605-003, 711-585-150 or 711-585-152, Jackson Immuno Research, ME, USA) for 4-6 hrs at RT. Images were acquired with an LSM 710 laser-scanning confocal microscope (Carl Zeiss, Germany), processed in Zen Black 2012 and analyzed in ImageJ. To minimize bias during quantification, matched regions across animals were selected for IHC according to GFP expression levels. To mitigate any cross-talk between distinct fluorescent channels, spectral unmixing was used and a regions of interest (ROI) around each mCherry+ or tagBFP+ cell was manual drawn and overlaid to the channel (Alexa Fluor 647) designed for FoxP2. The expression level of FoxP2 in each ROI was estimated by measuring the mean intensity value from the nucleus of each ROI and subtracting the background intensity. The expression level of FoxP2 in control cells are estimated from 20 random non-infected cells (FoxP2+mCherry-tagBFP-) from the same slice where the expression in the mCherry+ and tagBFP+ cells are measured.

Immunoblotting (IB) was carried out as described previously(Xiao et al., 2011). Cell lysates from each sample were separated by SDS-PAGE and transferred to an Immuno-Blot PVDF Membrane (162-0177, Bio-Rad Lab., CA, USA), then blocked with 1% skim milk in TBST (tris-buffered saline with 0.1% Tween-20) for 1 hr at RT. The membrane was incubated with primary antibodies overnight at 4°C, washed with TBST, and reacted with the appropriate horseradish peroxidase (HRP)-conjugated species-specific secondary antibodies (NA934 and NA931, Sigma-Aldrich, MO, USA; AP180P, Millipore, MA, USA) for 1 hr at RT. The signals were detected by Clarity western ECL substrate (170-5060, Bio-Rad Lab., CA, USA).

The primary antibodies used were: Goat anti-FOXP2(ab1307, Abcam, MA, USA), Goat anti-FOXP2(sc-21069, Santa Cruz Bio., TX, USA), mouse anti V5 tag(R960-25, Invitrogen, CA, USA), rabbit anti-RFP (mCherry, 600-401-379, Rockland, PA, USA), mouse anti-RFP (mCherry, 200-301-379, Rockland, PA, USA), rabbit anti-GFP rabbit (A11122, Invitrogen, CA, USA), chicken anti-GFP (AB16901, Millipore, MA, USA), rabbit anti-tRFP (tagBFP, AB233, Evrogen, Moscow, Russia) and mouse anti-GAPDH (MAB374, Millipore, MA, USA). The specificity of primary antibodies against FoxP2, RFP or GFP were confirmed by two independent primary antibodies for both IHC and IB.

Statistics

Shapiro-Wilk Test was performed for all behavioral data to test for normality of underlying distributions. Unless otherwise noted, statistical significance was tested with non-parametric statistical tests; Wilcoxon signed-rank tests and Wilcoxon rank-sum tests were used where appropriate. Kruskal-Wallis test was performed when comparisons were made across more than two conditions (e.g., baseline vs stimulation day vs recovery day) from individual

145 animal, whereas Friedman tests were performed when data was pooled across animals and comparisons were made
146 across more than two conditions. Statistical significance refers to *p < 0.05, **p < 0.02. Statistical details for all
147 experiments are included in their corresponding figure legends.
148
149 **Tissue Processing for single-nucleus RNA Sequencing (snRNA-seq)**
150
151 Adult zebra finches(120-140dph) were injected with CS-shFoxP2(n=2) and CS-shScr(n=2) constructs and scarified at
152 180-200dph. Birds were put down prior to lights in the to ensure that singing behavior did not affect our results. Each
153 bird was rapidly decapitated and its brain was placed in ice-cold ACSF (126 mM NaCl, 3 mM KCl, 1.25 mM NaH_2PO_4,
154 26 mM $NaHCO_3$, 10 mM D-(+)-glucose, 2 mM $MgSO_4$, 2 mM $CaCl_2$) bubbled with carbogen gas (95% O2, 5% CO2). The
155 cerebellum was removed with a razor blade and the cerebrum was glued to a specimen tube for sectioning with a VF-
156 200 Compresstome (Precisionary Instruments). Coronal 500 µm sections were made in ice-cold ACSF and allowed to
157 recover in room temperature ACSF for 5 min. Area X punches were placed into a tube containing ACSF on ice until all
158 punches were collected and pooled from 2 birds per condition. Tissue punches were dounce-homogenized in 500 µl
159 ice-cold Lysis Buffer (10 mM Tris pH 7.4, 10 mM NaCl, 3 mM $MgCl_2$, 0.1% IGEPAL CA-630) and transferred to a clean 2
160 ml tube. Then, 900 µl of 1.8 M Sucrose Cushion Solution (NUC201-1KT, Sigma, MO, USA) was added and pipette-
161 mixed with nuclei 10 times. 500 µl of 1.8 M Sucrose Cushion Solution was added to a second clean 2 ml tube, and the
162 nuclei sample was layered on top of the cushion without mixing. The sample was centrifuged at 13,000 × g for 45 min
163 at 4°C and all but ~100 µl of supernatant was discarded to preserve the pellet. The pellet was washed in 300 µl Nuclei
164 Suspension Buffer (NSB) (1% UltraPure BSA (AM2618, Thermo Fisher Scientific, MA, USA) and 0.2% RNase inhibitors
165 in PBS) and centrifuged at 550 × g for 5 min at 4°C. All but ~50 µl of supernatant was discarded and the pellet was
166 resuspended in the remaining liquid and filtered through a FLOWMI 40 µm tip strainer (H13680-0040, Bel-Art, NJ,
167 USA). Samples were diluted to 1000 nuclei/µl with NSB for targeting 10,000 nuclei for snRNA-seq. Libraries were
168 prepared using the Chromium Single Cell 3' Library & Gel Bead Kit v3 according to the manufacturer's instructions
169 and sequenced using an Illumina NovaSeq 6000 at the North Texas Genome Center at UT Arlington.
170
171 **Pre-processing of snRNA-seq Data**
172
173 Raw sequencing data was obtained as binary base cells (BCL files) from the sequencing core. 10X Genomics
174 CellRanger v.3.0.2 was used to demultiplex the BCL files using the *mkfastq* command. Extracted FASTQ files were
175 quality checked using FASTQC v0.11.5. Paired-end FASTQ files (26 bp long R1 – cell barcode and UMI sequence; 124
176 bp long R2 – transcript sequence) were then aligned to a reference Zebra Finch genome (bTaeGut1_v1.p) from UCSC
177 Genome Browser, and reads were counted as number of unique molecular identifiers (UMIs) per gene per cell using
178 10X Genomics CellRanger v.3.0.2 *count* command. Since the libraries generated are single-nuclei libraries, reported
179 UMIs per gene per cell accounts for reads aligned to both exons and introns. This was achieved by creating a
180 reference genome and annotation index for pre-mRNAs.
181
182 **snRNA-seq Clustering Analysis**
183
184 The resulting count matrices from the data pre-processing steps were analyzed with the Seurat analysis pipeline in R
185 (v.3.0, https://satijalab.org/seurat/v3.0/pbmc3k_tutorial.html). Cells with more than 10,000 UMI and more than 5%
186 mitochondrial genes were filtered out to exclude potential doublets and dead or degraded cells (Fig. S1). As
187 described in the Seurat pipeline, the data were log-normalized and scaled using a factor of 10,000 and regressed to
188 the covariates of the number of UMI and percent mitochondrial genes. Top variable genes were identified and
189 principal components (PCs) were calculated from the data. PCs to include were identified by "ElbowPlot" in Seurat,
190 where PCs are ranked according to the percentage of variance each one explains; PCs were excluded after the last
191 noticeable drop in explanatory power. With the selected PCs, the Louvain algorithm was then used to identify
192 clusters within the data. Clusters were visualized with uniform manifold approximation and projection (UMAP) in two
193 dimensions.
194
195 **snRNA-seq Dataset Integration and Differential Gene Expression**

In order to compare gene expression between CS-shScr+ and CS-shFoxP2+ birds, the datasets must first be combined. Each dataset was processed independently as described in Clustering Analysis up until the point of normalization. Once normalized, the datasets were integrated as described in the Seurat integrate pipeline. The integrated dataset was then regressed to the covariates of the number of UMI and percent mitochondrial genes. The subsequent analysis proceeded as described in Clustering Analysis. The integrated data were used to generate the integrated clusters and UMAP plot (Figure S3A), as recommended by the Seurat pipeline (https://satijalab.org/seurat/v3.0/immune_alignment.html). Comparisons of gene expression across clusters, and across datasets, were pulled from each individual RNA assay, as recommended by the Seurat pipeline. Differentially expressed genes were identified using the Wilcoxon ranked sum test with the Bonferroni correction as implemented in the Seurat pipeline. For a gene ontology analysis, the top differentially expressed genes in *FoxP2*+ cells between CS-shScr+ and CS-shFoxP2+ birds were loaded into ToppGene (https://toppgene.cchmc.org) with default settings and then gene family categories were sorted for similarity with REVIGO (http://revigo.irb.hr) on the "small (0.5)" similarity setting. For an analysis of differentially-expressed genes related to autism in *FoxP2*+ cells of CS-shFoxP2+ birds, genes were sorted based on scores in the Simons Foundation Autism Research Initiative (https://www.sfari.org), where Category 1 is "high confidence."

Data and Code Availability

The NCBI Gene Expression Omnibus (GEO) accession number for the snRNA-sequencing data in this manuscript is in submission. Codes for data pre-processing, clustering, and differential gene expression analysis are available at GitHub repository (in submission).

Cluster Function Assignment

Clusters were assigned functional identities based on the expression of gene markers established in the literature (see Table S1 below) as visualized in Figure 1E. Identity names were as specific as possible given the confidence and clarity of the gene expression pattern in relation to expression in other clusters. When gene markers have been established in Area X in zebra finches specifically in addition to mammals, this is noted in the table with a "ZF". For downstream analyses, specific UMAP coordinates were used to further specify MSNs (UMAP_1 > -5 & UMAP_2 > -6) and PNs (UMAP_1 > 0 & UMAP_2 > -7.5).

Table S1

Gene Symbol	Gene Name	Identity	Reference
Gad2	Glutamate decarboxylase 2	GABAergic	(Saunders et al., 2018)
Slc17a6	Solute carrier family 17 member 6	Glutamatergic	(Saunders et al., 2018)
Ppp1r1b	Protein phosphatase 1 regulatory subunit 1B (also known as DARPP-32)	MSN (ZF)	(Reiner et al., 2004; Saunders et al., 2018)
FoxP2	Forkhead box P2	MSN (ZF)	(Haesler et al., 2004; Mendoza et al., 2015)
FoxP1	Forkhead box P1	MSN (ZF)	(Mendoza et al., 2015)
Tac1	Tachykinin precursor 1	MSN (ZF)	(Carrillo and Doupe, 2004;

			Reiner et al., 2004)
Penk	Proenkephalin	PN (ZF)	(Carrillo and Doupe, 2004)
Pvalb	Parvalbumin	Interneuron (ZF)	(Reiner et al., 2004)
Sst	Somatostatin	Interneuron (ZF)	(Reiner et al., 2004)
Npy	Neuropeptide Y	Interneuron (ZF)	(Reiner et al., 2004)
Nos1	Nitric oxide synthase 1	Interneuron (ZF)	(Reiner et al., 2004)
Chat	Choline O-acetyltransferase	Interneuron (ZF)	(Carrillo and Doupe, 2004; Reiner et al., 2004)
Lrig1	Leucine rich repeats and immunoglobulin like domains 1	Astrocyte	(Zhang et al., 2014)
Mbp	Myelin basic protein	Oligodendrocyte	(Zhang et al., 2014)
Pdgfra	Platelet-derived growth factor receptor A	Oligodendrocyte Precursor	(Zhang et al., 2014)
Csf1r	Colony stimulating factor 1 receptor	Microglia	(Zhang et al., 2014)
Flt1	Vascular endothelial growth factor receptor 1	Endothelial	(Zhang et al., 2014)
Gene Symbol	Gene Name	Identity	Reference
Gad2	Glutamate decarboxylase 2	GABAergic	(Saunders et al., 2018)
Slc17a6	Solute carrier family 17 member 6	Glutamatergic	(Saunders et al., 2018)
Ppp1r1b	Protein phosphatase 1 regulatory subunit 1B (also known as DARPP-32)	MSN (ZF)	(Reiner et al., 2004; Saunders et al., 2018)
FoxP2	Forkhead box P2	MSN (ZF)	(Haesler et al., 2004; Mendoza et al., 2015)
FoxP1	Forkhead box P1	MSN (ZF)	(Mendoza et al., 2015)
Tac1	Tachykinin precursor 1	MSN (ZF)	(Carrillo and Doupe, 2004; Reiner et al., 2004)
Penk	Proenkephalin	PN (ZF)	(Carrillo and Doupe, 2004)

Pvalb	Parvalbumin	Interneuron (ZF)	(Reiner et al., 2004)
Sst	Somatostatin	Interneuron (ZF)	(Reiner et al., 2004)
Npy	Neuropeptide Y	Interneuron (ZF)	(Reiner et al., 2004)
Nos1	Nitric oxide synthase 1	Interneuron (ZF)	(Reiner et al., 2004)
Chat	Choline O-acetyltransferase	Interneuron (ZF)	(Carrillo and Doupe, 2004; Reiner et al., 2004)
Lrig1	Leucine rich repeats and immunoglobulin like domains 1	Astrocyte	(Zhang et al., 2014)
Mbp	Myelin basic protein	Oligodendrocyte	(Zhang et al., 2014)
Pdgfra	Platelet-derived growth factor receptor A	Oligodendrocyte Precursor	(Zhang et al., 2014)
Csf1r	Colony stimulating factor 1 receptor	Microglia	(Zhang et al., 2014)
Flt1	Vascular endothelial growth factor receptor 1	Endothelial	(Zhang et al., 2014)

229
230

Supplemental Figures 1-7

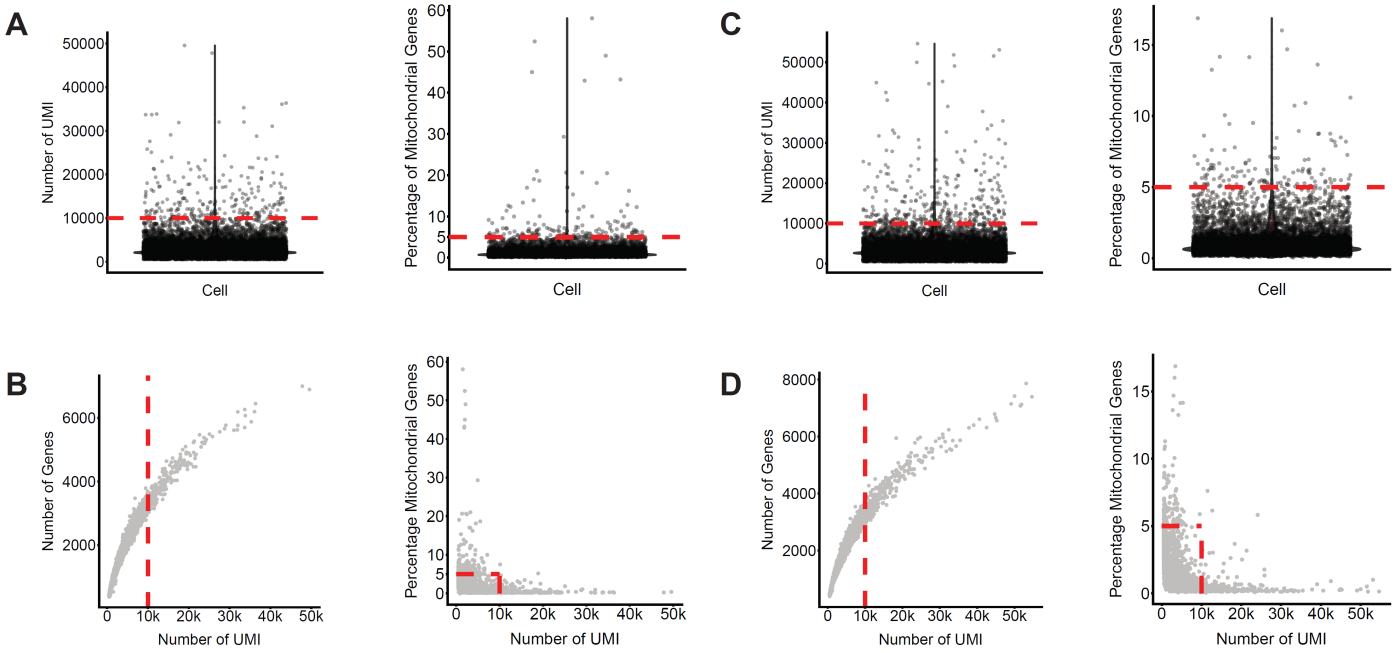
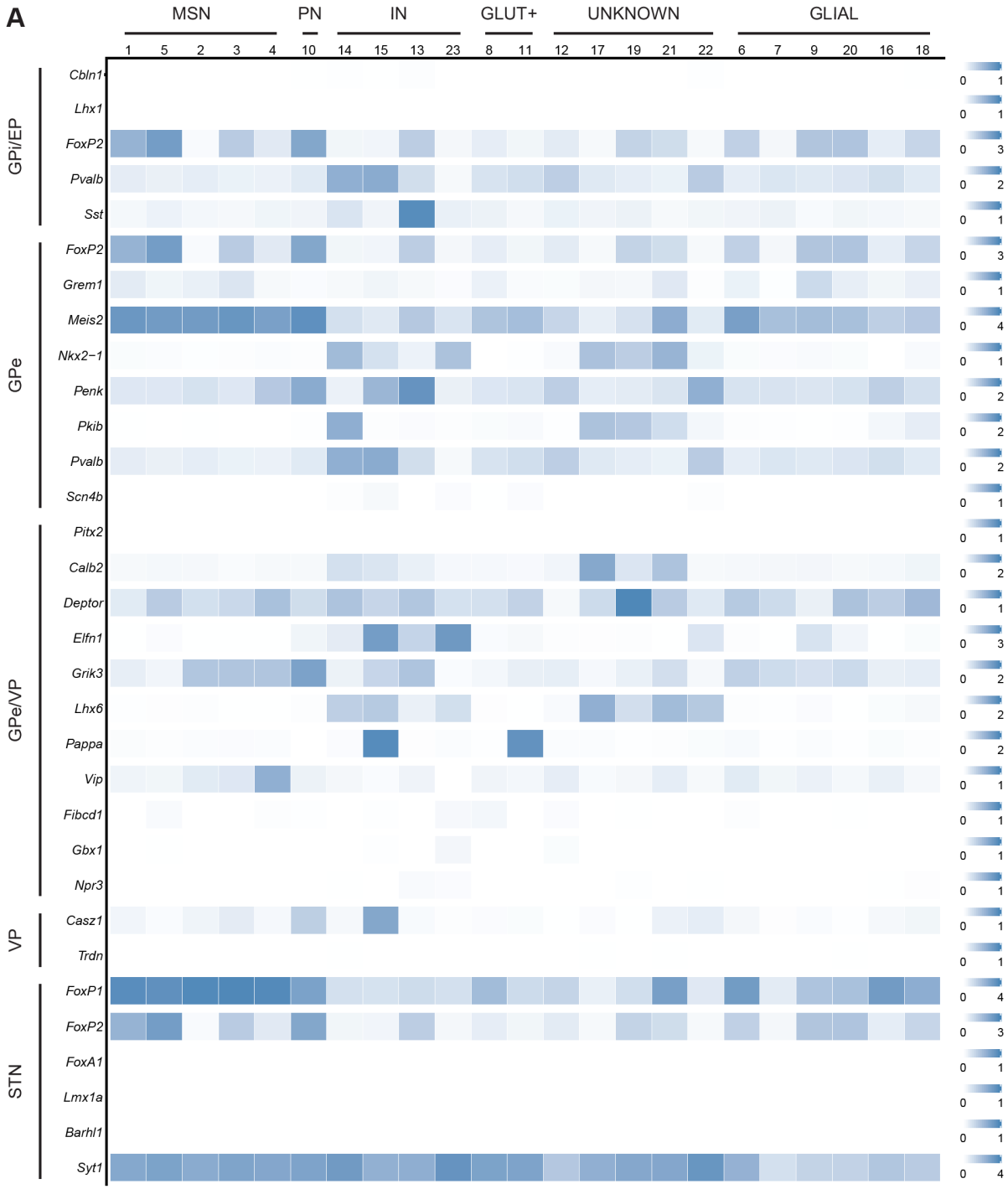
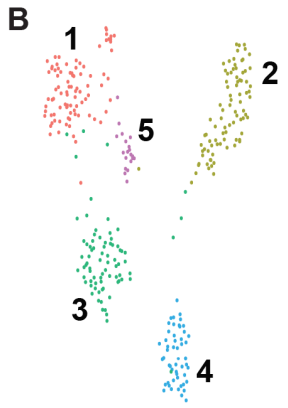
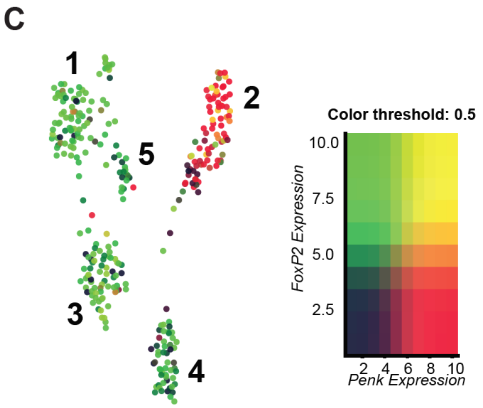


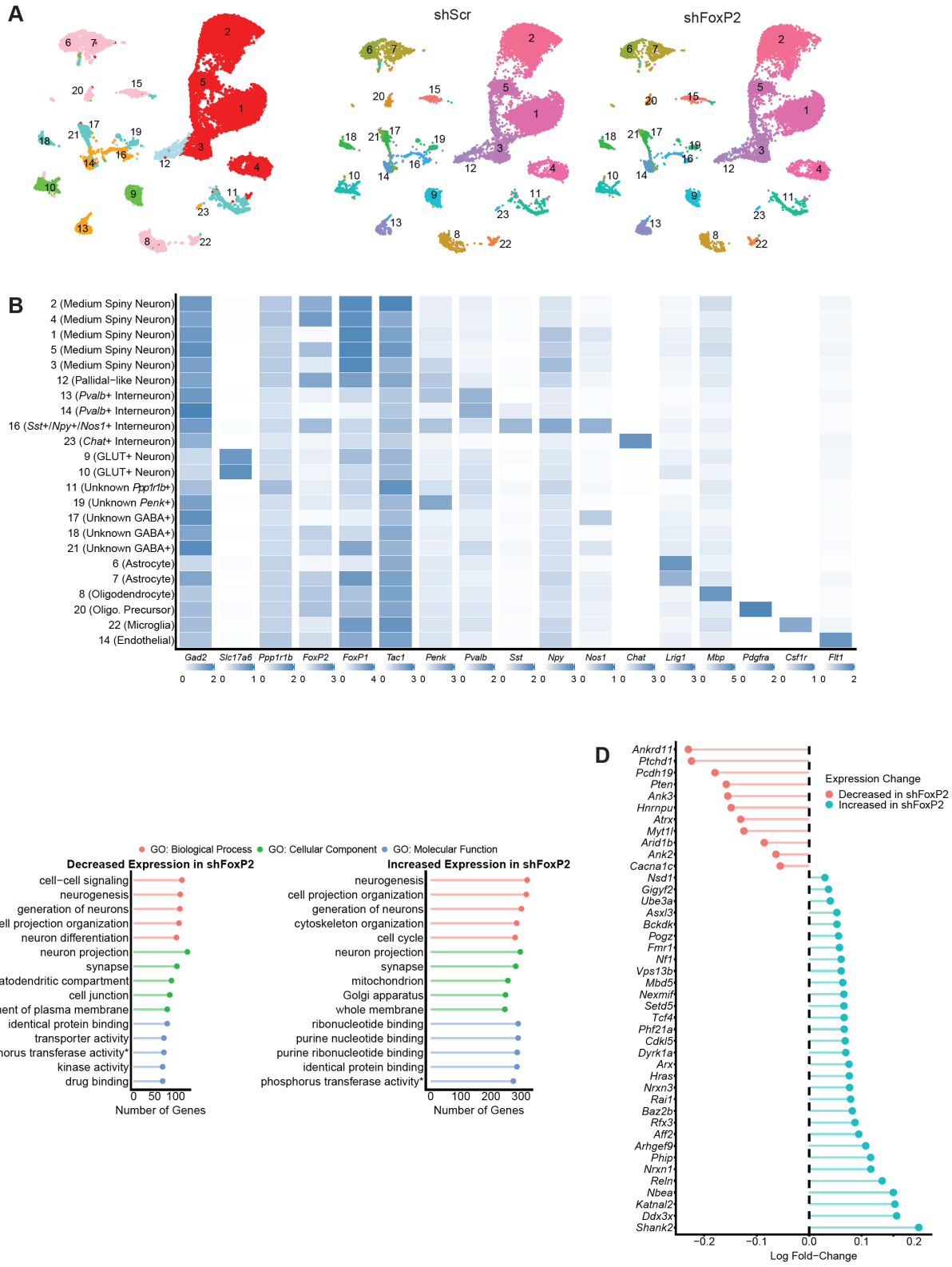
Figure S1 Details of single-cell transcriptomics.
(A) For CS-shScr, density plot of the number of UMIs per cell (left) and the percentage of mitochondrial genes in each cell (right). The analysis only included cells with UMI < 10,000 and < 5% mitochondrial genes (indicated by the red dashed line).
(B) For CS-shScr, a scatterplot of the number of UMI and the number of genes (left) or percentage of mitochondrial genes (right). Each dot is a cell. The cells within the red dashed box, corresponding to the filters in (A), were the cells analyzed.
(C) For CS-shFoxP2, density plot of the number of UMIs per cell (left) and the percentage of mitochondrial genes in each cell (right). The analysis only included cells with UMI < 10,000 and < 5% mitochondrial genes (indicated by the red dashed line).
(D) For CS-shFoxP2, a scatterplot of the number of UMI and the number of genes (left) or percentage of mitochondrial genes (right). Each dot is a cell. The cells within the red dashed box, corresponding to the filters in (C), were the cells analyzed.


A
MSN
PN
IN
GLUT+
UNKNOWN
GLIAL
1 5 2 3 4 10 14 15 13 23 8 11 12 17 19 21 22 6 7 9 20 16 18
GPi/EP
Cbln1
Lhx1
FoxP2
Pvalb
Sst
GPe
FoxP2
Grem1
Meis2
Nkx2−1
Penk
Pkib
Pvalb
Scn4b
GPe/VP
Pitx2
Calb2
Deptor
Elfn1
Grik3
Lhx6
Pappa
Vip
Fibcd1
Gbx1
Npr3
VP
Casz1
Trdn
STN
FoxP1
FoxP2
FoxA1
Lmx1a
Barhl1
Syt1

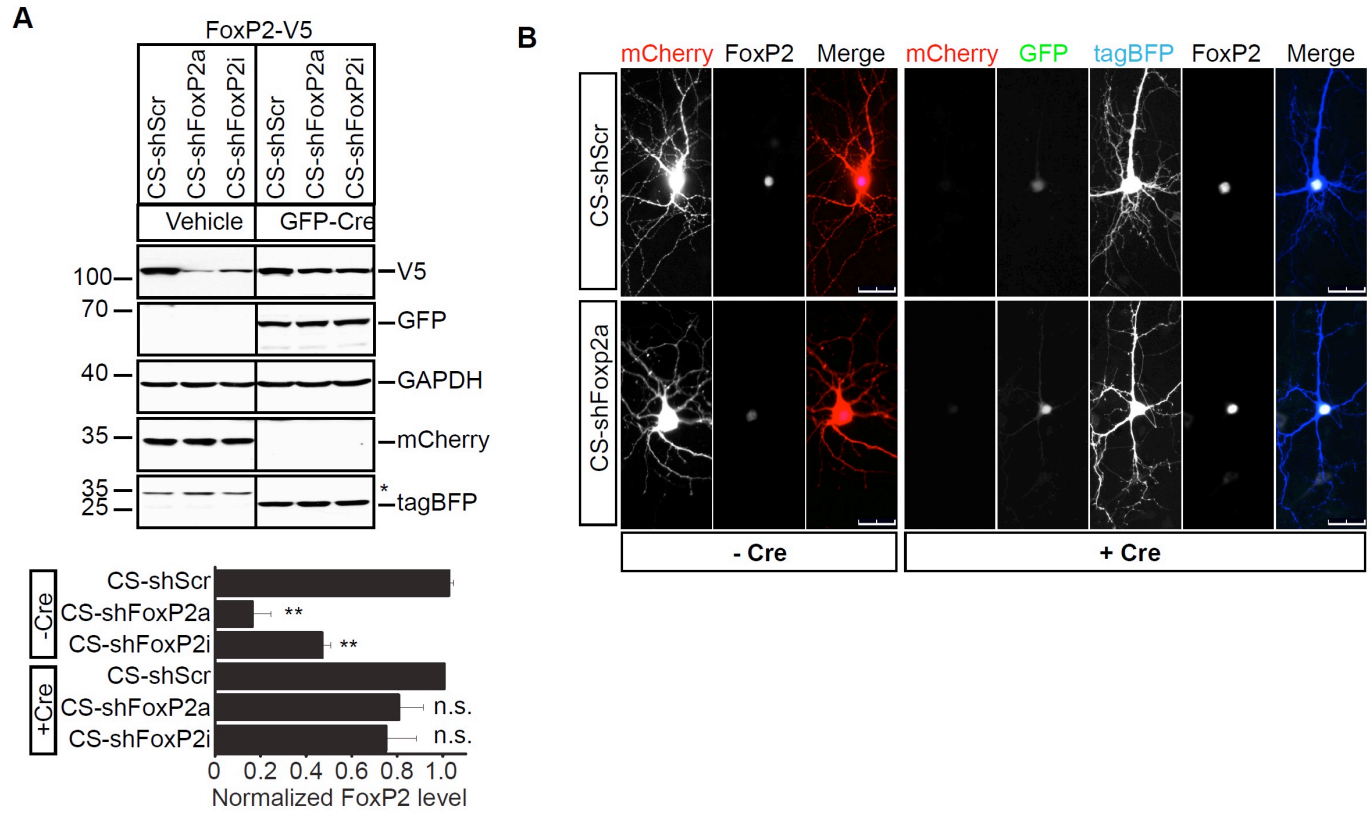

B
1
2
5
3
4

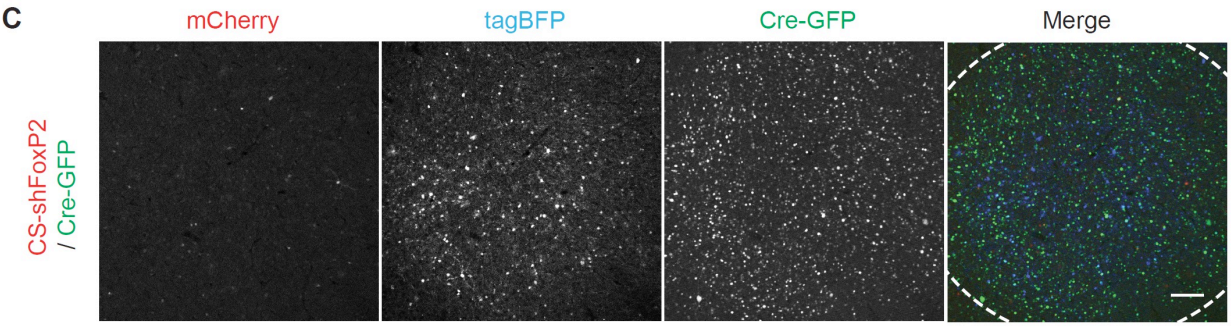

C
1
2
5
3
4
Color threshold: 0.5
FoxP2 Expression
10.0
7.5
5.0
2.5
2 4 6 8 10
Penk Expression

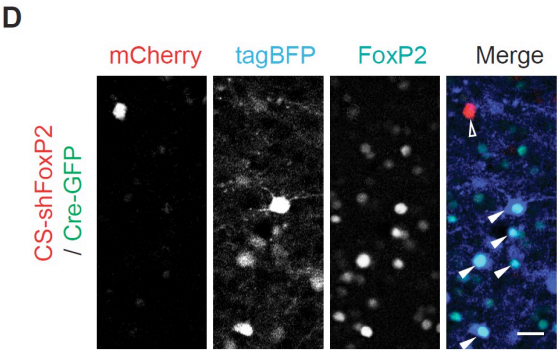
252 **Figure S2 Supplemental gene markers of basal ganglia regions.**
253 (A) As in Figure 1E, expression was normalized globally across all genes, but a different scale is shown for
254 each gene based on the highest normalized value. Gene markers were selected from published studies
255 (GPi/EP: (Saunders et al., 2018; Wallace et al., 2017); GPe: (Abdi et al., 2015; Saunders et al., 2018);
256 GPe/VP: (Saunders et al., 2018); VP: (Saunders et al., 2018); STN: (Papathanou et al., 2019; Saunders et al.,
257 2018)). No clear analogues of GPi cells, prototypical GPe cells, VP cells, or STN cells are seen in Area X.
258 (B) A UMAP projection of a sub-cluster analysis of PN from CS-shScr (325 nuclei total). Clusters are
259 numbered in ascending order by decreasing size (1-largest; 5-smallest).
260 (C) UMAP projection of PN clusters, taken from Panel (B), with each nucleus colored according the
261 normalized expression of *FoxP2* and *Penk*. Expression value colors are relative to each gene (0-lowest value
262 for that gene; 10-highest value for that gene).
263

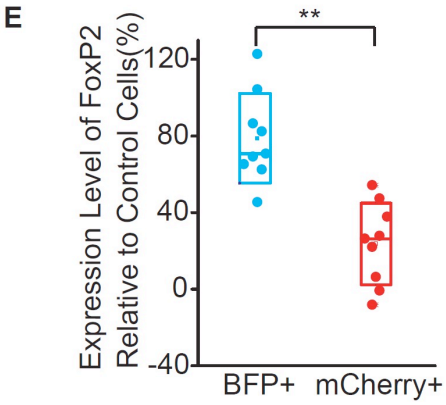

A
shScr
shFoxP2
B
2 (Medium Spiny Neuron)
4 (Medium Spiny Neuron)
1 (Medium Spiny Neuron)
5 (Medium Spiny Neuron)
3 (Medium Spiny Neuron)
12 (Pallidal−like Neuron)
13 (Pvalb+ Interneuron)
14 (Pvalb+ Interneuron)
16 (Sst+/Npy+/Nos1+ Interneuron)
23 (Chat+ Interneuron)
9 (GLUT+ Neuron)
10 (GLUT+ Neuron)
11 (Unknown Ppp1r1b+)
19 (Unknown Penk+)
17 (Unknown GABA+)
18 (Unknown GABA+)
21 (Unknown GABA+)
6 (Astrocyte)
7 (Astrocyte)
8 (Oligodendrocyte)
20 (Oligo. Precursor)
22 (Microglia)
14 (Endothelial)
Gad2
Slc17a6
Ppp1r1b
FoxP2
FoxP1
Tac1
Penk
Pvalb
Sst
Npy
Nos1
Chat
Lrig1
Mbp
Pdgfra
Csf1r
Flt1
C
GO: Biological Process
GO: Cellular Component
GO: Molecular Function
Decreased Expression in shFoxP2
cell−cell signaling
neurogenesis
generation of neurons
cell projection organization
neuron differentiation
neuron projection
synapse
somatodendritic compartment
cell junction
intrinsic component of plasma membrane
identical protein binding
transporter activity
phosphorus transferase activity*
kinase activity
drug binding
Number of Genes
Increased Expression in shFoxP2
neurogenesis
cell projection organization
generation of neurons
cytoskeleton organization
cell cycle
neuron projection
synapse
mitochondrion
Golgi apparatus
whole membrane
ribonucleotide binding
purine nucleotide binding
purine ribonucleotide binding
identical protein binding
phosphorus transferase activity*
Number of Genes
D
Expression Change
Decreased in shFoxP2
Increased in shFoxP2
Ankrd11
Ptchd1
Pcdh19
Pten
Ank3
Hnrnpu
Atrx
Myt1l
Arid1b
Ank2
Cacna1c
Nsd1
Gigyf2
Ube3a
Asxl3
Bckdk
Pogz
Fmr1
Nf1
Vps13b
Mbd5
Nexmif
Setd5
Tcf4
Phf21a
Cdkl5
Dyrk1a
Arx
Hras
Nrxn3
Rai1
Baz2b
Rfx3
Aff2
Arhgef9
Phip
Nrxn1
Reln
Nbea
Katnal2
Ddx3x
Shank2
Log Fold−Change

266 **Figure S3 Details of CS-shFoxP2 clustering and clustering of combined datasets.**
267 (A) Left, a UMAP projection of combined nuclei 27,245 (total) between CS-shScr and CS-shFoxP2. Middle
268 and right, cluster compositions split by dataset.
269 (B) A heatmap of normalized expression for genes used to assigned identities to cell types. Expression was
270 normalized globally across all genes, but a different scale is shown for each gene based on the highest
271 normalized value.
272 (C) The top five families of gene functions affected in CS-shFoxP2+ birds for each of the three main
273 categories of a gene ontology analysis (Biological Process, Cellular Component, and Molecular Function).
274 (D) Log fold-changes in expression of genes in CS-shFoxP2+ birds that are scored by the SFARI Gene
275 database as most strongly linked to autism.
276


A
FoxP2-V5
CS-shScr
CS-shFoxP2a
CS-shFoxP2i
CS-shScr
CS-shFoxP2a
CS-shFoxP2i
Vehicle
GFP-Cre
100
70
40
35
35
25
V5
GFP
GAPDH
mCherry
tagBFP
*
-Cre
+Cre
CS-shScr
CS-shFoxP2a
CS-shFoxP2i
CS-shScr
CS-shFoxP2a
CS-shFoxP2i
**
**
n.s.
n.s.
0 0.2 0.4 0.6 0.8 1.0
Normalized FoxP2 level
B
mCherry FoxP2 Merge mCherry GFP tagBFP FoxP2 Merge
CS-shScr
CS-shFoxp2a
- Cre
+ Cre


C
mCherry
tagBFP
Cre-GFP
Merge
CS-shFoxP2
/ Cre-GFP


D
mCherry
tagBFP
FoxP2
Merge
CS-shFoxP2
/ Cre-GFP


E
**
Expression Level of FoxP2
Relative to Control Cells(%)
120
80
40
0
-40
BFP+
mCherry+

277
278

Figure S4 Validation of CS constructs.

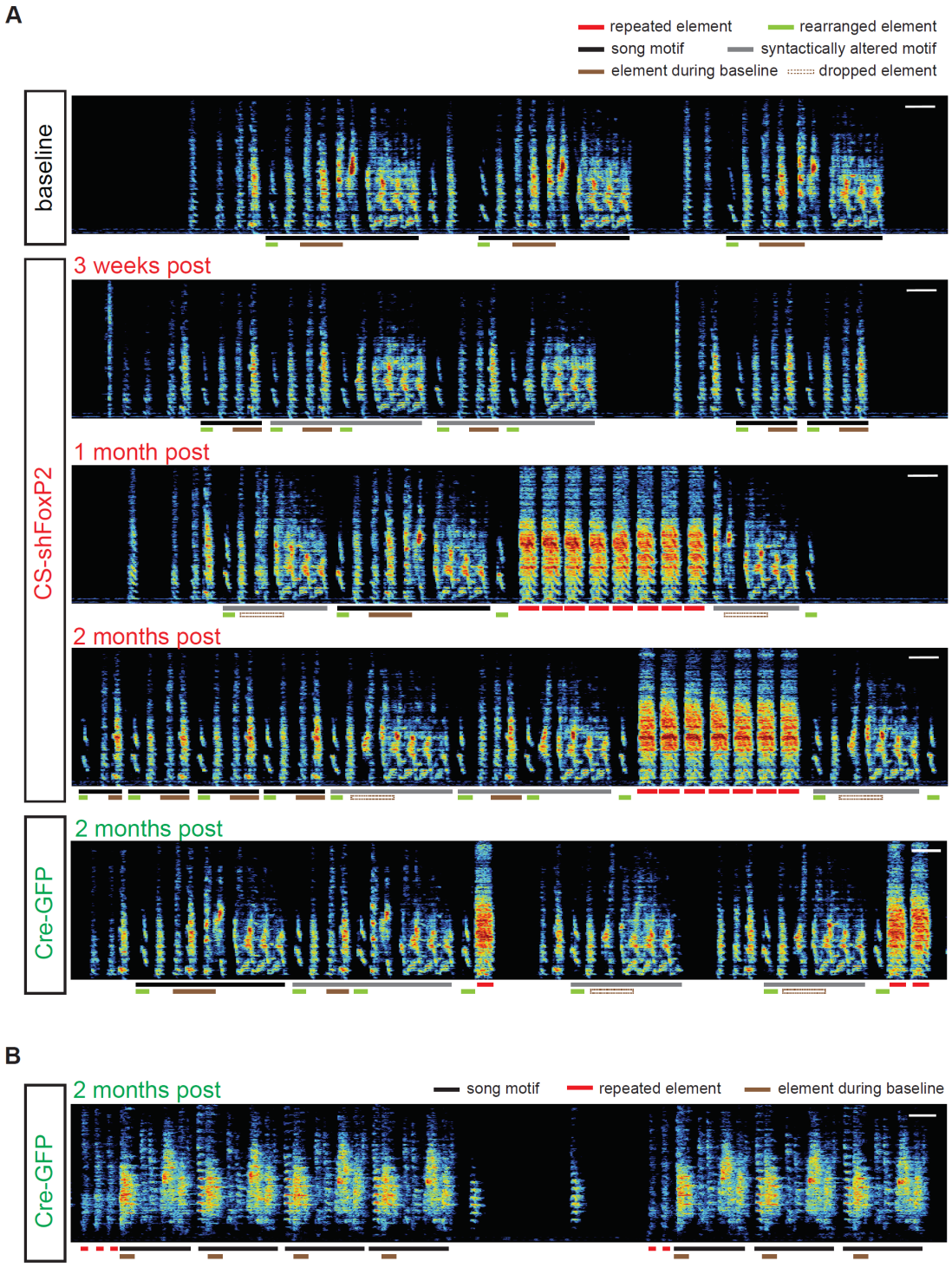
(A) Validation of CS constructs *in vitro* with western blotting. Two independent small hairpin RNAs against zebra finch FoxP2 gene (CS-shFoxP2a & i) and scrambled hairpin (CS-shScr) together with V5-tagged zebra finch FoxP2 were co-transfected into HEK stable cell lines expressing either vehicle or GFP-Cre as indicated. At 72 h after transfection, lysates of cells were subjected to immunoblotting with V5(to detect FoxP2), GFP (to detect Cre-GFP), GAPDH, RFP (to detect mCherry) and tagBFP antibody. The expression levels of FoxP2 were quantified from three independent experiments on the bottom panel. Both CS-shFoxP2a and CS-shFoxP2i constructs resulted in equivalent downregulation of FoxP2 protein levels ($p<0.0001$ and $p=0.0004$, ANOVA), which were rescued in the presence of Cre recombinase ($p=0.08$ and $p=0.93$, ANOVA).

(B) Validation of CS constructs in the primary culture. CS-shScr or -shFoxP2 together with V5-tagged zebra finch FoxP2 were co-transfected into mouse cortical primary culture with or without Cre-GFP (+Cre & -Cre respectively). For both the CS-shScr and -shFoxP2 constructs, the expression of mCherry was maintained in the absence of Cre, whereas the expression of mCherry was turned off and BFP was turned on in the presence of Cre. Scale bar, 30 µm.

(C) Representative parasagittal section shows the expression pattern of mCherry and tagBFP in Area X of an adult bird injected with CS-shFoxP2 and Cre-GFP constructs. Dashed lines outline the border of Area X. Scale bar, 100 µm.

(D) Enlarged confocal image shows that the expression of FoxP2 in mCherry+(open) and tagBFP+(filled) cells within Area X of an adult bird injected with CS-shFoxP2 and Cre-GFP constructs. Scale bar, 20 µm.

(E) Quantification of the expression level of FoxP2 in mCherry+ and tagBFP+ cells (red 23.7±16.4% vs blue 78±18% relative to control cells) within Area X of adult birds injected with CS-shFoxP2 and Cre-GFP constructs (n=9 slices from 3 birds). The expression level of FoxP2 in tagBFP+ cells is significantly higher than in mCherry+ cells ($p=0.0002$, Mann-Whitney test). Box indicates the median ± 1.0 SD, mean shown by open dot.


A
repeated element
rearranged element
song motif
syntactically altered motif
element during baseline
dropped element
baseline
CS-shFoxP2
3 weeks post
1 month post
2 months post
Cre-GFP
2 months post
B
Cre-GFP
2 months post
song motif
repeated element
element during baseline

303
304

Figure S5 Changes in song following knockdown and/or rescue of FoxP2 in Area X.

(A) Spectrograms of song recorded on the baseline day, 3 weeks, 1 month, and 2 months after injection of CS-shFoxP2 construct and 2 months after injection of Cre-GFP construct in Area X of an adult bird. A de novo syllable (red line) emerged in song bouts one-month post injection of the CS-shFoxP2 construct and the number of repetitions of this syllable was maintained for up to four months (data not shown) post injection of CS-shFoxP2 construct. In a subset of motifs, a portion of motif (brown and empty brown lines) was omitted one-month post injection of the CS-shFoxP2 construct. In another subset of motifs, the syllable immediately following that portion of the motif was replaced with the first syllable (green line) as early as three weeks post injection of the CS-shFoxP2 construct. The de novo syllable (red line) was retained in the end of a motif, whereas the number of repetitions in each song bout was significantly decreased 2 months following injection of Cre-GFP construct. All other changes in song caused by FoxP2 KD were not rescued 2 months following injection of Cre-GFP construct. Scale bar, 200 ms.

(B) Spectrograms of one bout of song recorded two months after bilateral injection of Cre-GFP construct in Area X of adult birds which were previously injected with CS-shFoxP2 construct for 2 months (previous spectrograms were illustrated in Figure 4E). The number of repetitions of introductory elements in each song bout was restored to baseline level, and previous omitted syllable was fully recovered 2 months following injection of Cre-GFP construct.

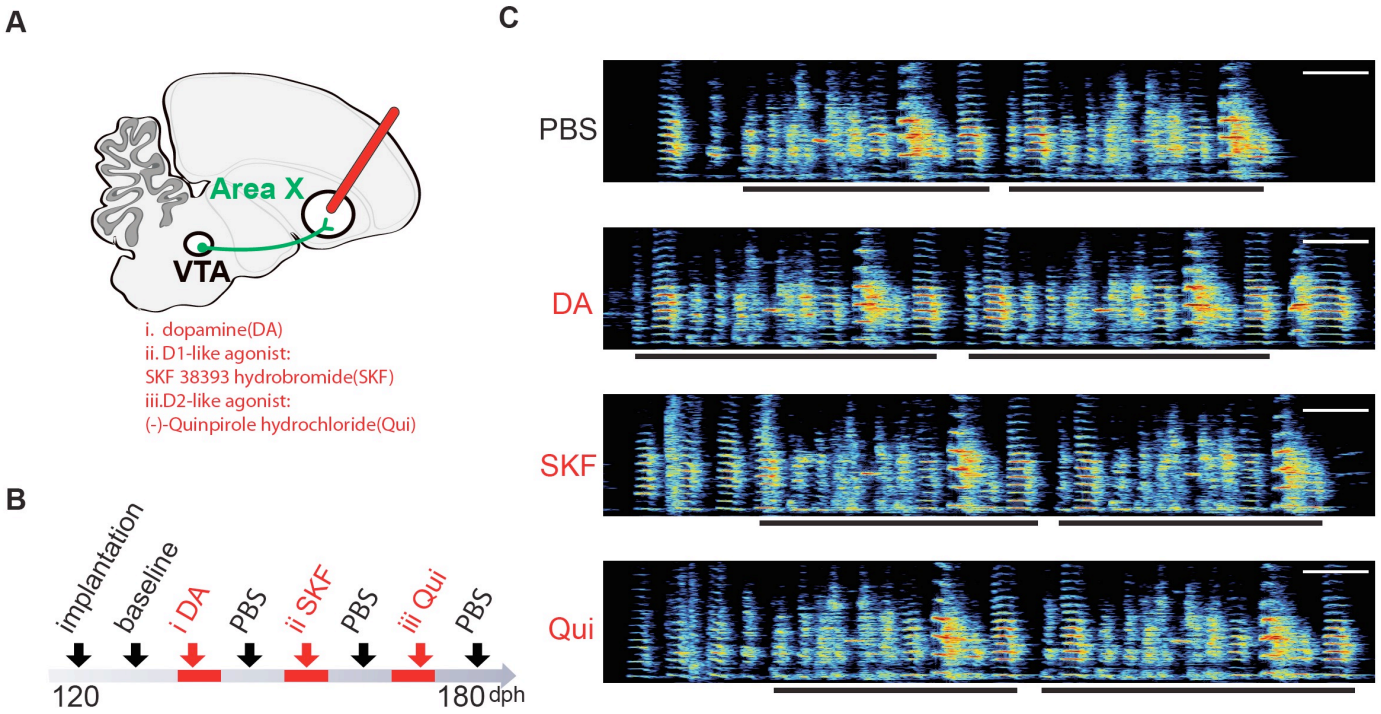


Figure S6 Lack of changes in vocal repetitions with pharmacological manipulation of dopamine or dopamine receptors.

(A-B) Schematic of experimental design. Dopamine (DA, i), D1-like agonist: SKF 38393 hydrobromide (SKF, ii.) and D2-like agonist: (-)-Quinpirole hydrochloride (Qui, iii.) were infused bilaterally into Area X of behaving adult birds through a microdialysis system. Each chemical was delivered individually for 4-8 days, a week apart from each other.

(C) Spectrograms of song recorded from one adult bird implanted with microdialysis probes in Area X on the baseline day (infused with PBS) and third day during chronic infusion of DA, SKF and Qui. Black lines indicate motifs. Scale bar, 200 ms.

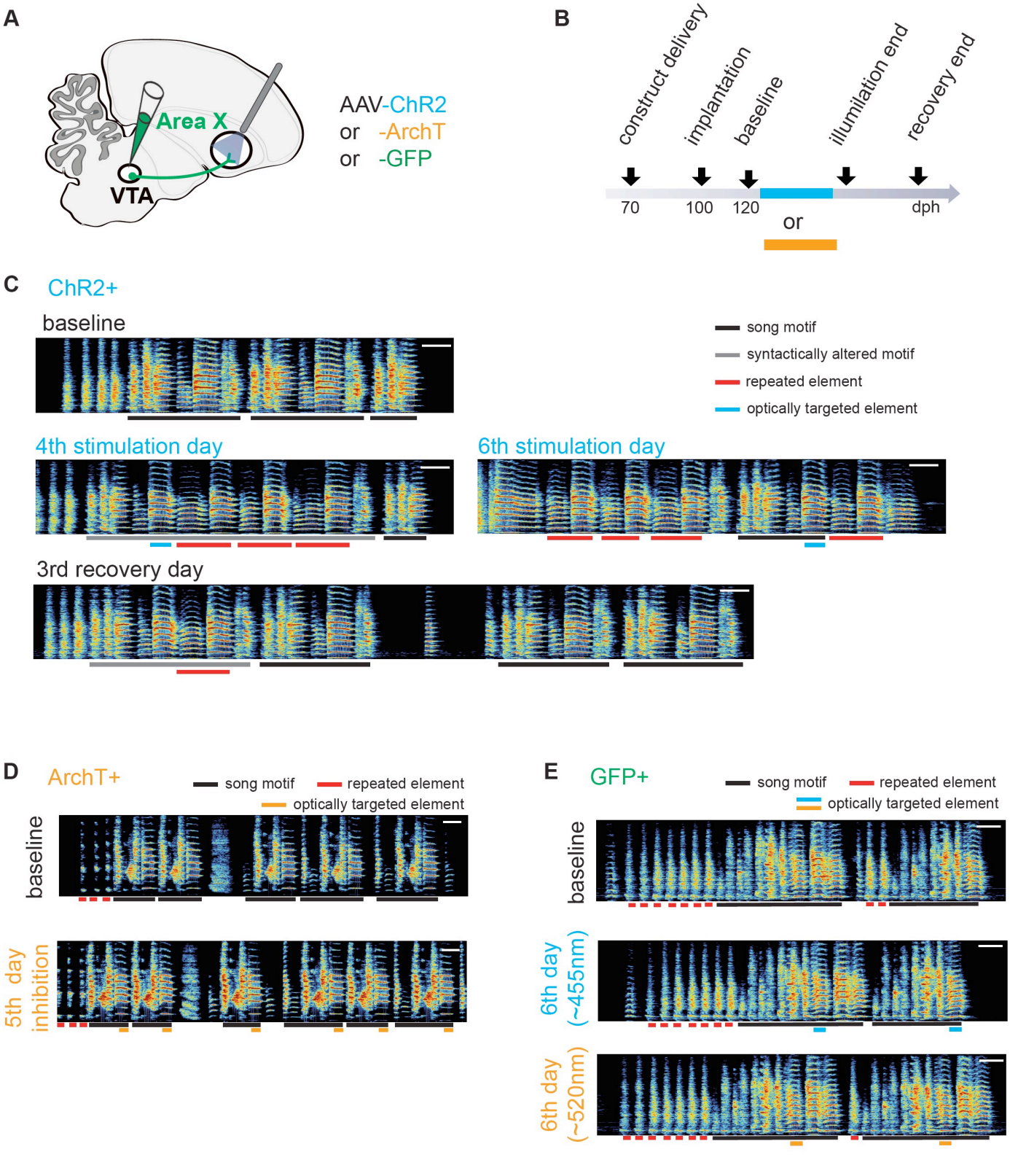

A
Area X
VTA
AAV-ChR2
or -ArchT
or -GFP
B
construct delivery
implantation
baseline
illumilation end
recovery end
70
100
120
dph
or
C
ChR2+
baseline
song motif
syntactically altered motif
repeated element
optically targeted element
4th stimulation day
6th stimulation day
3rd recovery day
D
ArchT+
song motif
repeated element
optically targeted element
baseline
5th day inhibition
E
GFP+
song motif
repeated element
optically targeted element
baseline
6th day (~455nm)
6th day (~520nm)

Figure S7 Song Contingent Optical Manipulation of Dopamine Terminals in Area X.
(A-B) Schematic of experimental design for optogenetic manipulation of dopamine release from VTA terminals. Optogenetic construct (ChR2 or ArchT) or GFP control construct was delivered into VTA in juvenile birds (~70 days post hatch(dph)) and cannula pointing to Area X were not bilaterally implanted until the song crystallized(~100dph). Optogenetic manipulations began at least 1week post cannulas were implanted to allow birds to fully recover and to begin singing again (blue box, wavelength of LED ~455 nm; green box, wavelength of LED ~520 nm). Songs recorded before the illumination started and after the illumination ceased were referred as song of baseline and recovery, respectively.
(C) Spectrograms of song bout recorded from a ChR2+ bird on the baseline day, 4th/6th stimulation day and 3rd recovery day illustrating a bird repeating a pair of syllables within the song motif or in the beginning or end of the song motif. Scale bar, 200 ms.
(D-E) Lack of changes in vocal repetitions in birds expressing ArchT or GFP following optical manipulations. (D) Spectrograms of song recorded from a ArchT+ bird on the baseline day and 5th inhibition day. Light pulses (~520 nm, 100 ms) were delivered over the target syllable in a subset of variants during inhibition days. Scale bar, 200 ms. (E) Spectrograms of song recorded from a GFP+ bird on the baseline day(top) and 6th day of optical illumination. Light pulses (middle, blue light, ~455nm, 100 ms; bottom, green light, ~520 nm, 100 ms) were delivered over the target syllable (blue or orange line) in a subset of variants during illumination days. Scale bar, 200 ms.

References:

Abdi, A., Mallet, N., Mohamed, F.Y., Sharott, A., Dodson, P.D., Nakamura, K.C., Suri, S., Avery, S.V., Larvin, J.T., Garas, F.N., *et al.* (2015). Prototypic and arkypallidal neurons in the dopamine-intact external globus pallidus. J Neurosci *35*, 6667.
Ali, F., Otchy, T.M., Pehlevan, C., Fantana, A.L., Burak, Y., and Ölveczky, B.P. (2013). The basal ganglia is necessary for learning spectral, but not temporal, features of birdsong. Neuron *80*, 494-506.
Carrillo, G.D., and Doupe, A.J. (2004). Is the songbird Area X striatal, pallidal, or both? An anatomical study. J Comp Neurol *473*, 415-437.
Haesler, S., Rochefort, C., Georgi, B., Licznerski, P., Osten, P., and Scharff, C. (2007). Incomplete and Inaccurate Vocal Imitation after Knockdown of FoxP2 in Songbird Basal Ganglia Nucleus Area X. PLoS Biology *5*, e321.
Haesler, S., Wada, K., Nshdejan, A., Morrisey, E.E., Lints, T., Jarvis, E.D., and Scharff, C. (2004). *FoxP2* expression in avian vocal learners and non-learners. J Neurosci *24*, 3164.
Hamaguchi, K., and Mooney, R. (2012). Recurrent Interactions between the Input and Output of a Songbird Cortico-Basal Ganglia Pathway Are Implicated in Vocal Sequence Variability. Journal of Neuroscience *32*, 11671-11687.
Kao, M.H., and Brainard, M.S. (2006). Lesions of an avian basal ganglia circuit prevent context-dependent changes to song variability. J Neurophysiol *96*, 1441-1455.
Mendoza, E., Tokarev, K., Düring, D.N., Retamosa, E.C., Weiss, M., Arpenik, N., and Scharff, C. (2015). Differential coexpression of FoxP1, FoxP2, and FoxP4 in the zebra finch (*Taeniopygia guttata*) song system. J Comp Neurol *523*, 1318-1340.
Miller, M.N., Cheung, C.Y.J., and Brainard, M.S. (2017). Vocal learning promotes patterned inhibitory connectivity. Nat Commun *8*, 2105-2105.

384 Papathanou, M., Björklund, Å.K., Martis-Thiele, M.M., and Wallén-Mackenzie, Å. (2019). Single-nuclei
385 transcriptomic analysis of the subthalamic nucleus reveals different Pitx2-positive subpopulations.
386 bioRxiv, 677724.
387 Reiner, A., Laverghetta, A.V., Meade, C.A., Cuthbertson, S.L., and Bottjer, S.W. (2004). An
388 immunohistochemical and pathway tracing study of the striatopallidal organization of Area X in the
389 male zebra finch. J Comp Neurol *469*, 239-261.
390 Roberts, T.F., Gobes, S.M., Murugan, M., Olveczky, B.P., and Mooney, R. (2012). Motor circuits are
391 required to encode a sensory model for imitative learning. Nat Neurosci *15*, 1454-1459.
392 Sakata, J.T., Hampton, C.M., and Brainard, M.S. (2008). Social modulation of sequence and syllable
393 variability in adult birdsong. J Neurophysiol *99*, 1700-1711.
394 Saunders, A., Johnson, C.A., and Sabatini, B.L. (2012). Novel recombinant adeno-associated viruses for
395 Cre activated and inactivated transgene expression in neurons. Frontiers in neural circuits *6*, 47.
396 Saunders, A., Macosko, E.Z., Wysoker, A., Goldman, M., Krienen, F.M., de Rivera, H., Bien, E., Baum, M.,
397 Bortolin, L., Wang, S.*, et al.* (2018). Molecular diversity and specializations among the cells of the adult
398 mouse brain. Cell *174*, 1015-1030.e1016.
399 Tchernichovski, O., Nottebohm, F., Ho, C.E., Pesaran, B., and Mitra, P.P. (2000). A procedure for an
400 automated measurement of song similarity. Anim Behav *59*, 1167-1176.
401 Tian, L.Y., and Brainard, M.S. (2017). Discrete Circuits Support Generalized versus Context-Specific
402 Vocal Learning in the Songbird. Neuron *96*, 1168-1177 e1165.
403 Wallace, M.L., Saunders, A., Huang, K.W., Philson, A.C., Goldman, M., Macosko, E.Z., McCarroll, S.A., and
404 Sabatini, B.L. (2017). Genetically distinct parallel pathways in the entopeduncular nucleus for limbic
405 and sensorimotor output of the basal ganglia. Neuron *94*, 138-152.e135.
406 Xiao, L., Chattree, G., Oscos, F.G., Cao, M., Wanat, M.J., and Roberts, T.F. (2018). A Basal Ganglia Circuit
407 Sufficient to Guide Birdsong Learning. Neuron *98*, 208-221 e205.
408 Xiao, L., Chen, D., Hu, P., Wu, J., Liu, W., Zhao, Y., Cao, M., Fang, Y., Bi, W., Zheng, Z.*, et al.* (2011). The c-
409 Abl-MST1 signaling pathway mediates oxidative stress-induced neuronal cell death. J Neurosci *31*,
410 9611-9619.
411 Zhang, Y., Chen, K., Sloan, S.A., Bennett, M.L., Scholze, A.R., Keeffe, S., Phatnani, H.P., Guarnieri, P.,
412 Caneda, C., Ruderisch, N.*, et al.* (2014). An RNA-sequencing transcriptome and splicing database of glia,
413 neurons, and vascular cells of the cerebral cortex. J Neurosci *34*, 11929.
414