Cortico-basal ganglia projecting neurons are required for juvenile vocal learning but not for adult vocal plasticity in songbirds

Miguel Sánchez-Valpuesta[1], Yumeno Suzuki[1], Yukino Shibata[1], Noriyuki Toji[2], Yu Ji[1], Nasiba Afrin[1], Chinweike Norman Asogwa[1], Ippei Kojima[1], Daisuke Mizuguchi[3], Satoshi Kojima[3], Kazuo Okanoya[4], Haruo Okado[5], Kenta Kobayashi[6], and Kazuhiro Wada[1, 2, 7]*

[1] Graduate School of Life Science, [2]Faculty of Science, and [7]Department of Biological Sciences, Hokkaido University, Sapporo, Hokkaido, 060-0810, Japan.
[3] Department of Structure & Function of Neural Networks, Korea Brain Research Institute (KBRI), Daegu, 41068, South Korea
[4] Department of Cognitive and Behavioral Sciences, The University of Tokyo, Meguro, Tokyo, 153- 8902, Japan.
[5] Tokyo Metropolitan Institute of Medical Science, Setagaya, Tokyo, 156-8506, Japan.
[6] National Institute for Physiological Sciences, Okazaki, Aichi, 444-8585, Japan.

*Corresponding author
Kazuhiro Wada MD, PhD
Associate Professor,
Faculty of Science, Hokkaido University
Room 910, Building No. 5, North 10, West 8, Kita-ku
Sapporo, Hokkaido 060-0810, Japan
E-mail: wada@sci.hokudai.ac.jp

Key words
Critical period, sensorimotor learning, time-locked firing, zebra finch, sensory feedback, behavioral fluctuation

Abstract

Birdsong, like human speech, consists of a sequence of temporally precise movements acquired through vocal learning. The learning of such sequential vocalizations depends on the neural function of the motor cortex and basal ganglia. However, it is unknown how the connections between cortical and basal ganglia components contribute to vocal motor skill learning, as mammalian motor cortices serve multiple types of motor action and most experimentally tractable animals do not exhibit vocal learning. Here, we leveraged the zebra finch, a songbird, as an animal model to explore the function of the connectivity between cortex-like (HVC) and basal ganglia (Area X), connected by $HVC_{(X)}$ projection neurons with temporally precise firing during singing. By specifically ablating $HVC_{(X)}$ neurons, juvenile zebra finches failed to copy tutored syllable acoustics and developed temporally unstable songs with less sequence consistency. In contrast, $HVC_{(X)}$-ablated adults did not alter their learned song structure, but generated acoustic fluctuations and responded to auditory feedback disruption by the introduction of song deterioration, as did normal adults. These results indicate that the cortico-basal ganglia input is important for learning the acoustic and temporal aspects of song structure, but not for generating vocal fluctuations that contribute to the maintenance of an already learned vocal pattern.

Significance

We addressed the question, “How do cortico-basal ganglia projecting neurons contribute to vocal learning?”. We performed specific ablation of the vocal cortical neurons projecting to the basal ganglia, $HVC_{(X)}$ neurons in a songbird, which generate temporally precise firing during singing. Specific ablation of $HVC_{(X)}$ neurons in juveniles caused deficits in learning the tutor song’s acoustics and less consistency of song sequence. In contrast, adult $HVC_{(X)}$ neuron ablation did not affect the degree of vocal fluctuations or cause alteration in song structure by auditory feedback inhibition. These results support the hypothesis that $HVC_{(X)}$ neurons are a neural substrate for transferring temporal signals, but not for regulating vocal fluctuations or conveying auditory feedback, to the basal ganglia for vocal learning and maintenance.

Introduction

Complex motor skills are composed of a series of sequential movements acquired through learning with repetitive practice (1, 2). Neural activity coding for temporal information is considered to play an important role in the coordination of motor exploration and performance evaluation during the learning and execution of sequential movements (3, 4). General temporal information for externally reinforced motor sequence learning, such as start or stop timing, has been shown to be transferred from the cortical areas involved in cognitive control to the basal ganglia (5-7). In addition, premotor and motor cortical areas have the potential to carry into the basal ganglia fine-grained temporal information more suited for the precise control of motor learning (8-10). Indeed, cortical-basal ganglia synaptic plasticity is necessary for the acquisition of motor sequences (11), implying a potential link between impairments of connectivity from the cortex to basal ganglia and motor control pathologies (12, 13). However, how the cortical-basal ganglia connection functionally and causally contributes to learning and maintenance of sequential motor skills remains largely unknown.

Birdsong is produced as a sequence of skilled vocal movements, which are acquired through vocal learning (14-16). Songbirds memorize and copy the acoustic and sequential features of a tutor's song during a critical period of vocal learning (**Figure 1A**). In the songbird brain, a distinct group of brain nuclei, called the song system, contributes to song learning and production (17, 18) (**Figure 1B**). The song system is composed of two major circuits: the posterior vocal motor pathway and the anterior forebrain pathway (AFP). Although the vocal motor pathway participates in song production (19-21), the AFP, which is homologous to the mammalian cortical-basal ganglia-thalamic loop, plays a crucial role in vocal motor learning (22-24).

Song nucleus HVC (used as a proper name) in nidopallium, which is analogous to the mammalian premotor cortex, stands on top of the hierarchy of the song system and projects to both the vocal motor pathway and the AFP (25). HVC is a critical site for both the production and learning of song (26-28) and contains two major subpopulations of projection neurons: $HVC_{(RA)}$ neurons that project to the nucleus robustus of the arcopallium (RA), which is the telencephalic output locus connecting to the tracheosyringeal part of the hypoglossal nucleus (nXIIts) (29, 30), and $HVC_{(X)}$ neurons projecting to the basal ganglia nucleus Area X in the AFP (31, 32). Both types of projection neurons are active at specific time points during singing renditions (8, 20, 33). It has been proposed that the firing of $HVC_{(X)}$ neurons collectively represents temporal sequence information during song renditions, but does not convey the properties of constitutive vocal gestures nor sensory feedback signals (8, 9, 34-38). Although the ablation of $HVC_{(X)}$ neurons in adults does not affect the execution of learned vocalization (39), the potential functional contribution of $HVC_{(X)}$ neurons in vocal learning remains unclear. In addition, the AFP is a crucial neural site for the generation of vocal exploration and the refinement of vocal performance using auditory feedback information (40-43). However, it is unknown how the temporally precise firing inputs from $HVC_{(X)}$ neurons to Area X relate to regulation of vocal variability and auditory-dependent song maintenance.

Here, we performed cell-type-specific ablation of $HVC_{(X)}$ neurons ($HVC_{(X)}$ ablation) to disrupt the transfer of sequential and temporally precise firing from a cortical-like region to the basal ganglia. To elucidate the cellular functions of $HVC_{(X)}$ neurons on song learning and maintenance, we ablated $HVC_{(X)}$ neurons in juvenile zebra finches before the initiation of vocal motor learning and analyzed their acquired songs. In addition, we examined the effect of $HVC_{(X)}$ ablation on the regulation of vocal variability and change in song structure after deprivation of

auditory feedback in adults.

Results

$HVC_{(X)}$ neuron-specific ablation in zebra finches

To achieve sufficient and specific ablation of $HVC_{(X)}$ neurons *in vivo*, we used self-complementary adeno-associated virus (scAAV) vectors to ensure a stronger and faster induction than normal single-stranded AAV genomes (44). We used AAV serotype 9 capsid (AAV9), which allows retrograde transport from the site of injection, and a Cre/FLEx switch system for the conditional induction of gene expression. We first injected scAAV9-Cre and scAAV9-FLEx-mRuby2 into Area X and HVC, respectively, to test the induction rate and timing of expression of a targeted gene (i.e., mRuby2) in $HVC_{(X)}$ neurons. One week after the injections, we observed the selective expression of mRuby2 protein in $HVC_{(X)}$ neurons (**Figure 1C and Supplementary figure 1A**). Neurotensin (NTS) mRNA was used as a marker of $HVC_{(X)}$ neurons. We confirmed that NTS mRNA was specifically expressed in $HVC_{(X)}$ neurons labeled with retrograde cholera toxin B (CTB) injected in Area X (n = 4; 96.9% ± 2.0% of total $HVC_{(X)}$ neurons), but not in $HVC_{(RA)}$ neurons (1.7% ± 0.7%) (**Figure 1D**). Using NTS labeling, a reliable estimation of residual $HVC_{(X)}$ neuron number after ablation could be performed without additional retrograde labeling from Area X to HVC. We then evaluated the ablation efficiency of $HVC_{(X)}$ neurons by the dual induction of diphtheria toxin A (dtA) (45, 46) and constitutively active caspase 3 (caCasp) (47) into the same cells using the Cre/FLEx switch system. Both dtA and caspase 3 have been shown to synergistically potentiate caspase 3-dependent apoptotic cell death (48, 49). For this purpose, we injected scAAV9-Cre into Area X and a mixture of scAAV9-FLEx-dtA and -caCasp into HVC in a test hemisphere. In the other hemisphere of the same animal, scAAV9-Cre and scAAV9-FLEx-mRuby2 were injected into Area X and HVC, respectively, to serve as the control hemisphere. The number of $HVC_{(X)}$

neurons was compared between the control and $HVC_{(X)}$-ablated hemispheres (**Figure 1D and Supplementary figure 1B**). As a result, the residual number of $HVC_{(X)}$ neurons was reduced to 23.9–57.2% (mean ± SD = 38.8% ± 13.5%) in the ablated hemisphere when compared against the control HVC of the same individual (n = 6; $p < 0.001$, unpaired *t* test). Although the cell ablation procedure did not achieve complete removal of $HVC_{(X)}$ neurons, our method using a mixture of dtA and caCasp showed a similar or higher effective reduction of target cells compared with previous efforts at cell ablation in songbirds (39, 42, 50).

Deficits in song learning and development by ablation of $HVC_{(X)}$ neurons

To examine the cell-type specific function of $HVC_{(X)}$ neurons in song learning and development, we bilaterally injected scAAV9-Cre and scAAV9-FLEx-dtA/-caCasp into Area X and HVC, respectively, to ablate $HVC_{(X)}$ neurons before the initiation of sensorimotor learning. The injected juveniles were tutored using playback songs [post-hatching day (phd) ~35] (**Figure 2A**). We continuously recorded their songs daily and later evaluated the residual number of $HVC_{(X)}$ neurons in the adult stage (phd 180–200) by fluorescence *in situ* hybridization (FISH) using an NTS probe. We used birds possessing $HVC_{(X)}$ neuron densities lower than 130 NTS+ cells/mm^2 in both hemispheres for further analyses [NTS+ cells/mm^2 in HVC (mean ± SD); $HVC_{(X)}$ ablation: 104.1 ± 10.8, control injection: 386.2 ± 33.7] (**Supplementary figure 2**). The degree of $HVC_{(X)}$ ablation in individual birds ranged from 68.6% to 76.3% (mean ± SD, 73.0% ± 2.8%) in $HVC_{(X)}$-ablated birds when compared with the average density of NTS+ cells in the HVC of control birds. $HVC_{(X)}$-ablated birds developed their songs from subsongs with unstable syllable acoustics into a more stable and consistent

song structure through the critical period of song acquisition. The timing of song stabilization of $HVC_{(X)}$-ablated birds tended to be delayed compared with control birds for both syllable acoustic and sequential features (**Figure 2B**). Thus, we used the syllable acoustics and sequences of acquired songs as behavioral parameters to evaluate the successfulness of song learning.

$HVC_{(X)}$-ablated birds showed deficits in copying the acoustic features of syllables from the tutored songs (**Figure 3A** and **Supplementary figure 3**). Although there was no significant difference in the number of unique syllables between control and $HVC_{(X)}$-ablated birds in adults (phd 150–160) ($p = 0.12$, Student's t test), $HVC_{(X)}$-ablated birds showed a significant decrease in the average of each syllable similarity score toward the original tutored syllables compared with those of control birds (**$p < 0.01$, Student's t test) (**Figure 3B**). $HVC_{(X)}$-ablated birds did not completely fail to mimic the syllables of the tutored song. Rather, they copied some populations of syllables from the tutor song, although other population of acquired syllables did not belong to the tutor song (**Figure 3A**). This mixture of copied and non-copied syllables caused a significantly higher coefficient of variation (CV) in the syllable similarity scores in the $HVC_{(X)}$-ablated birds versus control birds (*$p < 0.05$, Student's t test) (**Figure 3B**). We further tested whether there was an association between the existing number of $HVC_{(X)}$ neurons and the syllable similarity score of individual birds (**Figure 3C**). We founf a significant correlation between the two factors ($p < 0.027$, $r = 0.758$, Pearson's correlation coefficient), indicating that the potential contribution of $HVC_{(X)}$ neurons to the accurate learning of syllable acoustics.

We further noticed that two of five $HVC_{(X)}$-ablated birds produced acoustically unstable

syllables with variable entropy variances and durations even in a mature adult stage (phd >150) (**Figure 3D**). In addition, although there was no significant difference in the median of inter-syllable gap duration between control and $HVC_{(X)}$-ablated birds' groups, two of the $HVC_{(X)}$-ablated birds produced strikingly short inter-syllable gaps (median < 25 ms) (**Figure 3E, F**). A certain number of $HVC_{(X)}$-ablated birds had defects not only in song learning but also in producing stable acoustic structures of song.

Ablation of $HVC_{(X)}$ neurons in juveniles increased the instability of the adult song syllable sequence

We next examined the developmental effects of $HVC_{(X)}$ ablation on the syllable sequence in songs. We used the syllable similarity matrix (SSM) method, which allowed quantitative analysis of the frequency of characteristic syllable transitions in songs without using human-biased procedures for syllable identification (51). In this analysis, two successive paired- and repetitive-syllable transitions were respectively measured as motif or repetitive indices (**Figure 4A** and **Supplementary figure 4**; see **Materials and methods**). The analysis indicated that $HVC_{(X)}$-ablated birds displayed a significant decrease in the frequency of paired-syllable transitions, forming motif structures when compared with control birds (*$p < 0.05$, Student's t test) (**Figure 4B**). Two of the $HVC_{(X)}$-ablated birds produced a relatively higher degree of repetitive-syllable transitions in their songs compared with the songs of control birds, despite the existence of a variety of individual differences among the ablated birds (**Figure 4B**). In addition, we calculated song consistency to examine the sequence variability of their songs (23). In agreement with SSM analysis, $HVC_{(X)}$-ablated birds showed significantly decreased song consistency when compared with control birds (*$p < 0.05$, Student's t test) (**Figure 4C**).

Furthermore, the scores for song consistency showed a large variation among the $HVC_{(X)}$-ablated birds, which is reflected as a large degree in the coefficient of variation (CV) in song consistency. Taken together, these results indicated that the reduction of $HVC_{(X)}$ neurons not only caused deficits in copying tutor song acoustics but also led to the abnormal development of song phonology and sequence.

Ablation of $HVC_{(X)}$ neurons in adults did not alter crystallized song structure or vocal fluctuation

As found in a previous study, chromophore-laser ablation of $HVC_{(X)}$ neurons in the adult stage does not change the learned song structure (39), as with the lack of apparent alterations in syllable acoustics and sequence following adult Area X lesions (23, 24). However, Area X lesions alter the within-syllable variability in fundamental frequency (FF) and cause a tangent effect on cross-rendition variability in syllable FF even in the adult stage, suggesting that Area X activity is related to the role of the AFP in generating exploratory vocal variability (43, 52, 53). However, it remains unknown whether $HVC_{(X)}$ ablation also influences vocal variability in a similar manner to Area X lesions.

To examine this possibility, we injected scAAV9-Cre and scAAV9-FLEx-dtA/-caCasp into Area X and HVC, respectively, of adult zebra finches (phd <120), causing a similar degree of ablation, ranging from 68.3% to 86.1% (mean ± SD, 79.1% ± 8.3%), as observed in $HVC_{(X)}$ ablation in juveniles (**Supplementary figure 2**). Owing to the absence of any changes in song structure at a few days after virus injection (i.e., $HVC_{(X)}$ ablation had yet to occur yet due to the time lag of gene induction by AAVs), we considered that bilateral AAV injections did not cause direct physical damage (**Figure 5A, B**). For quantitative comparisons of song structure changes

between pre-ablation and 2−3 weeks post-ablation, we used the motif index based on the SSM for the syllable sequence (51), Kullback–Leibler (K–L) distance based on two-dimensional syllable scatter plots for syllable acoustics (duration and mean FM) (54, 55) (**Figure 5C**), and motif duration for song tempo (43, 56). Similar to a previous study using the laser ablation of adult $HVC_{(X)}$ neurons (39), our $HVC_{(X)}$ ablation in the adult stage did not induce changes in these parameters of song structure (**Figure 5D–F**), indicating learning state-dependent effects of $HVC_{(X)}$ ablation on the production of structured songs. We then examined the potential contribution of $HVC_{(X)}$ neurons to the generation of song variability by focusing on "within-syllable variability" and "cross-rendition variability" in syllable FF between control and $HVC_{(X)}$-ablated adult birds. We found no obvious alterations in pre/post changes in both within- and cross-rendition syllable variability in FF between control and $HVC_{(X)}$-ablated birds (**Figure 5G, H**). In line with this finding, there were no significant differences in both the within- and cross-rendition variability in FF of the syllables between the pre- and post-injection states of $HVC_{(X)}$-ablated birds (**Supplementary figure 5**). These results indicate that $HVC_{(X)}$-ablated birds generate vocal fluctuations to a similar degree compared with the control birds.

Auditory feedback-dependent song changes after $HVC_{(X)}$ ablation

Although we found a consistent generation of vocal fluctuations after adult $HVC_{(X)}$ ablation, such $HVC_{(X)}$-ablated birds should transmit deteriorated temporal information to the AFP. Therefore, we hypothesized that $HVC_{(X)}$ ablation may have a different effect on auditory feedback-dependent vocal plasticity than Area X lesions. The neural activity output from the AFP to the song nucleus RA plays an important role in the regulation of vocal fluctuations (43,

57, 58), which could, in turn, be a driving force for the induction of song degradation after the disruption of auditory feedback (52, 57-59). Therefore, to examine whether adult $HVC_{(X)}$-ablated birds undergo a degradation of song structure by auditory deprivation, we prepared adult $HVC_{(X)}$-ablated and deafened birds by bilateral cochlear extirpation after 3 weeks after virus injection (**Figure 6A**). Ablation procedures caused a similar degree of $HVC_{(X)}$ ablation, ranging from 66.2% to 78.8% (mean ± SD, 72.2% ± 5.3%), as observed in the juveniles and adults without deafening (**Supplementary figure 2**). We then compared the degree of song degradation after deafening with age-matched deafened-alone birds. We found that deafened birds after $HVC_{(X)}$ ablation had a similar trajectory and variation of degradation of both sequence and acoustic features as the degree of degradation shown in deafened-alone birds (**Figure 6B–D)**. A comparison of the motif indices and K–L distances indicated significant differences between the pre- and post-deafening time points in both $HVC_{(X)}$-ablated and deafened and control deafened-alone birds (**Figure 6C, D**). However, there were no significant differences in the motif indices and K–L distances between the two groups after 1 and 2 months. These results indicate that AFP output activity generated under a severely reduced number of $HVC_{(X)}$ neurons is still sufficient to induce auditory-dependent song structural change.

Discussion

We utilized songbirds as a model system to investigate the cell-type-specific function of cortical neurons projecting to the basal ganglia on motor skill learning, motor fluctuation, and sensory feedback dependent alterations of learned motor skills. The song system, which includes the AFP, is a discrete neural circuit that shares a number of similarities with mammalian motor circuits (17, 18, 60). Unlike mammalian motor circuits, the song system is specialized for a well-defined behavior, singing, which therefore allows us to investigate cell-type specific functions in the circuits through quantitative behavioral measurements. Using AAV-induced genetic ablation of $HVC_{(X)}$ neurons in juvenile and adult zebra finches, we found a functional contribution of $HVC_{(X)}$ neurons to learning the acoustic and temporal aspects of song structure. In contrast, we did not observe the effects of $HVC_{(X)}$ ablation on the generation of vocal fluctuations and auditory-dependent changes in already learned songs. These results broadly support the hypothesis that the temporally precise activity of $HVC_{(X)}$ neurons is crucial for vocal motor learning, but is not involved in the generation of vocal fluctuation or transfer sensory feedback signals to the basal ganglia nucleus Area X (3, 34, 36-38).

The ablation of $HVC_{(X)}$ neurons indicated similarities and differences to lesions of the basal ganglia song nucleus Area X (23, 52). Both $HVC_{(X)}$ ablation and Area X lesions occurring before the initiation of vocal motor learning induced similar effects on song acquisition and execution. Both lesions affect the ability to copy tutor songs and lead to the production of sequentially unstable songs with inconsistency of syllable and inter-syllable gap durations, which are different from the effects of LMAN (lateral magnocellular nucleus of the anterior nidopallium) lesion (22-24). In addition, neither $HVC_{(X)}$ neurons nor Area X are required for the rendition of learned structured songs (23, 39). However, auditory feedback-driven song

changes in adults were strikingly different between two cases. Like LMAN lesions, Area X lesions prevent deafening-induced song degradation (52, 59). In contrast, deafened birds after $HVC_{(X)}$ ablation showed a very similar trajectory of song structural changes compared with $HVC_{(X)}$-intact deafened birds at both the phonological and sequential levels (**Figure 6**). The degree of $HVC_{(X)}$ ablation in adult deafened birds was 66–78% in this study. In contrast, Area X lesions ranged from 75% to 100% in a previous study (52). Therefore, it is important to consider the possibility that the difference in enabling auditory feedback-dependent vocal plasticity between Area X lesions and those in our study is not caused by the irrelevance of the connection between HVC and Area X, but may reflect the qualitative and quantitative differences in the magnitude of disruption of Area X activity between the two studies. A previous electrophysiological study demonstrated that lesions of Area X diminished song-locked burst firing tendency in LMAN, but did not affect the firing rate during singing (52), suggesting that the generation and transmission of temporally biased burst firing signals from LMAN to RA is a crucial factor for the induction of auditory feedback-dependent song plasticity. In this scenario, $HVC_{(X)}$-ablated birds may show a dampened song-locked firing patterns but may maintain burst firing in LMAN, which still induces vocal variability via the AFP outflow to RA during singing. If so, our results may support the hypothesis that functional AFP-driven vocal fluctuation is generated by Area X, DLM (dorsal lateral nucleus of the medial thalamus), and LMAN independently from $HVC_{(X)}$-derived temporal information. In addition, our results support recent studies showing that auditory feedback signals are not transferred into the AFP via $HVC_{(X)}$ neurons (37, 38), but rather from other areas such as the ventral tegmental area, a region implicated in reinforcement learning (61, 62). However, we cannot rule out the possibility that other HVC cell populations may be a locus for the transfer of auditory feedback

signal from the auditory forebrain to the song system.

In general, virus injection-based cell ablation is technically limited in terms of the removal efficiency of targeted cells, often not achieving complete ablation (in this study, ~ 85% ablation was the maximum efficiency) when compared with transgenesis-based cell ablation. However, this technique still has benefits for exploring the function of $HVC_{(X)}$ neurons. In the zebra finch, individual $HVC_{(X)}$ neurons generate temporally precise sparse and brief bursts of spikes during each song rendition (8), with bursts of different $HVC_{(X)}$ neurons being generated at different time points in the song motif, covering both syllables and inter-syllable gaps. In addition, the axon terminal arborization of $HVC_{(X)}$ neurons in Area X lacks topographical organization between the two song nuclei, indicating that each $HVC_{(X)}$ neuron projects to most of Area X (63). These issues suggest that, during singing, cell assemblies of $HVC_{(X)}$ neurons transmit information about the successive current song-locked time to Area X as a continuous-temporal code that allows temporal specificity for song learning (3). Thus, the virus injection-based $HVC_{(X)}$-ablated birds can be thought of as a model system for motor learning displaying deteriorated internal temporal firing information. Therefore, in the ablated birds left with significantly decreased numbers of $HVC_{(X)}$ neurons, Area X should receive a temporally incomplete (but not completely extinct) code as a sequentially inconsistent time representation for the learning process. This may be one of the reasons why the ablated juveniles still retained the ability to copy a few syllables from the tutor songs and developed relatively structured songs, despite showing phonological and sequential instability (**Figure 3**), instead of producing completely unstructured songs. It is necessary to point out a potential abnormality in the HVC microcircuit resulting from depletion of an $HVC_{(X)}$ subpopulation. Although we did not examine the cell number of other HVC neuron subpopulations, such as $HVC_{(RA)}$ neurons and

interneurons, a previous study indicated that ablation of $HVC_{(X)}$ neurons in juveniles upregulates the incorporation of new $HVC_{(RA)}$ neurons in HVC, although $HVC_{(X)}$-ablated adults do not increase $HVC_{(RA)}$ neurons (39). The increased number of $HVC_{(RA)}$ neurons in $HVC_{(X)}$-ablated birds in the juvenile stage might cause the deficit in song learning and development through induction of an imbalance in synaptic connections between HVC and LMAN to RA and/or potential circuitry disruption within HVC. The same applies to other described functions of $HVC_{(X)}$ neurons in the HVC nucleus, such as retinoic acid synthesis (64). Here, the mRNA of its synthesizing enzyme is only expressed by $HVC_{(X)}$ neurons, but its proteins are found in neighboring $HVC_{(RA)}$ neurons (65) or the guidance of newly born $HVC_{(RA)}$ neurons by interactions with $HVC_{(X)}$ neurons (66). All these hypothesized functions of $HVC_{(X)}$ neurons in the HVC microcircuit would be affected by $HVC_{(X)}$ neuron ablation.

We showed the different learning state-dependent effects of $HVC_{(X)}$ ablation on song production, finding no apparent effects of $HVC_{(X)}$ neurons on the execution and maintenance of learned songs in the zebra finch, consistent with a previous study (39). However, owing to incomplete $HVC_{(X)}$ ablation, it is necessary to consider the possibility that the residual population of $HVC_{(X)}$ neurons could still fulfill their role in generating song fluctuation and regulating auditory-dependent song deterioration in adults. Hence, adult birds with ablated $HVC_{(X)}$ neurons could execute and maintain a learned song in a similar way to the control birds. Therefore, it remains crucial to perform transgenesis-based $HVC_{(X)}$ neuron-specific ablation, although transgenic songbirds have yet to become a widely used experimental approach

In summary, our results shed light on how cortical neurons projecting to the basal ganglia contribute to motor skill learning, thus confirming the importance of the inputs from the song motor nucleus HVC input to the AFP in song learning. Furthermore, our data portray the

learning state-specific role of cortical-basal ganglia projection neurons for vocal skill learning. Considering the important role of the cortical-basal ganglia circuit for speech learning and production in humans and the paucity of animal models for vocal learning (17, 18, 67-69), studies of the cortical- basal ganglia circuit in songbirds may enhance ourr understanding of the neural basis of vocal developmental and vocal communication disorders in humans.

Materials and methods

Ethics statement

All animal experiments were performed according to the guidelines of the Committee on Animal Experiments of Hokkaido University from whom permission for this study was obtained (Approval No. 13-0061). The guidelines are based on national regulations for animal welfare in Japan (Law for the Humane Treatment and Management of Animals; after partial amendment No. 105, 2011). For brain sampling, the birds were humanely killed by decapitation after being injected with a lethal dose of pentobarbital.

Animals

Male zebra finches were obtained from our breeding colony at Hokkaido University. The photoperiod was constantly maintained at a 13/11 h light/dark cycle with food and water provided *ad libitum*. The sex of the birds was checked by PCR to select male juveniles before experimental manipulation. Birds for the song developmental study were raised in individual breeding cages with their parents and siblings until phd 5–15. Juveniles (along with their siblings) were then raised in a sound-attenuation box by their mother with their siblings after removal of their father (removed before phd 15) until they could feed themselves (phd ~35). Juvenile birds were subsequently separated from their mother and siblings and housed in individual isolation boxes for song playback, with the same tutor song being played back from phd 30 to all developing juveniles until at least phd 140. Birds for the adult experiments were raised in individual breeding cages with their parents and siblings until phd 60–100 and then housed in common cages with other male birds.

Song recording, tutoring, and analysis

Songs were recorded using a unidirectional microphone (SM57, Shure, IL) connected to a computer with the sound event triggered by recording software Sound Analysis Pro (sap v2011.089; http://soundanalysispro.com/) (70). Each song bout was saved as a sound file (.wav file), including time information. Low frequency noise (< 0.5 kHz) and mechanical noise were filtered out using Avisoft-SASLab's (Avisoft Bioacoustics, Glienicke, Germany) band pass filter. With respect to song tutoring, birds were individually housed in a sound-attenuating box containing a mirror to reduce social isolation. Tutor songs were played five times in the morning and five times in the afternoon at 55–75 dB from a speaker (SRS-M30, SONY) passively controlled by Sound Analysis Pro.

For the analysis of similarity between pupil and tutor songs, the comparison of tutor and pupil syllable acoustic features was performed using the SAP program's similarity module. The score was calculated using the "symmetric" and "time courses" comparison settings after manually adjusting the thresholds for every syllable. Overall, 80–120 syllables, including multiple syllable types, were compared with syllables from tutor songs to obtain each similarity score between syllables in the pupil and tutor songs. The mean values of the similarity score for each syllable type in pupil songs against each syllable type in tutor songs were calculated, and the highest mean values were used as the similarity scores of each syllable type. We used the total mean value of the similarity scores of all syllable types for each individual bird. For the coefficient of variation (CV) of syllable similarity (shown in Figure 3B), the CV using the similarity scores of each syllable type was calculated for individual birds.

To analyze the syllable transitions, song similarity matrix (SSM) analysis was performed (51). In all, 250 syllables from songs chosen randomly at phd 150 were used. Introductory notes in a song were not included in the song rendition. A total of 100 serially separated ".son"-converted syllable files were transferred to the Avisoft CORRELATOR program to calculate the similarity scores between the syllables of two songs by the round-robin comparison. The score was calculated as the peak correlation coefficient between two syllables according to the following formula:

$$\Phi_{XY} = \frac{\sum_X \sum_Y ((a_{xy} - m_a) * (b_{xy} - m_b))}{\sqrt{\sum_X \sum_Y (a_{xy} - m_a)^2 * \sum_X \sum_Y (b_{xy} - m_b)^2}}$$

where m_a and m_b are the mean values of the spectrograms a and b, respectively. a_{xy} and b_{xy} are the intensities of the spectrogram points at the locations x and y, respectively. The syllable similarity score is a value ranging from 0 to 1. Similarity scores between the syllables in two song renditions were exported into cells in the Microsoft Excel spreadsheet by maintaining the syllable sequence order in the original songs. The spreadsheet was named with the information of the similarity scores between the syllables as an SSM. In this study, 10 SSMs per bird were prepared by the round-robin comparison of 250 syllables. To qualitatively visualize the information of syllable temporal sequences in songs, each cell in the SSM was color coded according to the value of the similarity score. A similarity score of 0.595 was used as the threshold to distinguish similar or different syllables. For the quantitative analysis of syllable temporal structures, the occurrence rate of characteristic patterns of binarized "2 row × 2 column" cells in the SSMs was calculated. For the binarized patterning of 2 × 2 cells in the SSMs, the R software program was used to find the most similar binarized pattern for each 2 × 2 cell in the SSM from 12 possible patterns. The "motif" pattern was defined as a "paired-

syllables transition," indicating the existence of two successive syllables that were different but with the same sequential order in two songs. This can be illustrated by "song 1 [··A B·····] vs. song 2 [····A B····]" (in this case, A and B represent two different syllables). The "repetition" pattern was a case of the existence of the "repetitive-syllables transition" by two successive identical or very similar syllables in two songs: for instance, "song 1 [·····A A··] vs. song 2 [···A A·····]." The mean of the occurrence rate of the motif and repetition patterns and their coefficients of variation (CV) from 10 total SSMs per an individual animal were used for statistical analyses.

For song sequence analysis, song consistency was measured (23). Sequence consistency is calculated as the sum of typical transitions per bout divided by the sum of total transitions per bout. This measures how consistently the bird sings the same transitions over several bouts. Syllable identification was performed and aligned by two different researchers without information on individual birds. For highly variable syllables, on the basis of acoustic morphology on the spectrogram and sequential position between singing rendition, we categorized them as an identical syllables.

To measure the dynamics of syllable acoustic changes between two time points, we quantified changes in syllable acoustic features and syllable populations as two-dimensional scatter density plots. Syllable duration (ms; denoted as m) and mean FM (denoted as n) were measured to calculate the Kullback–Leibler (K–L) distance (54, 55), which was adapted as a way to measure the distance between two sets of syllable populations by comparing their probability density distributions. Syllable segmentation was performed manually for all syllables on a SASLab spectrogram after turning the amplitude intensity to the maximum in order to clarify any continuities/discontinuities in syllable boundaries. In all, 150 syllables were used to generate a two-dimensional density scatter plot. The probability density functions of each set of syllables were estimated at two different time points a and b, as Qa and Qb for the two time points, and the K–L distance score was then calculated to compare the density functions. If we let qa(m, n) and qb(m, n) denote the estimated probabilities for bin (m = 20, n = 5) for time points a and b, respectively, then the K–L distance between Qa and Qb is defined as follows:

$$D_{KL}(Q_a||Q_b) = \sum_{m=a}^{M} \sum_{n=a}^{N} q_a\,(m,n) \log_2 \frac{q_a(m,n)}{q_b(m,n)}.$$

A larger value for the K–L distance corresponds to a lower similarity between the distributions of two sets of syllable populations at different time points. Thus, a K–L distance of 0 indicates a perfect match between two sets of syllable populations. These behavioral analyses were performed as blind, without information of the residual number of $HVC_{(X)}$ neurons of each

individual.

Effects of $HVC_{(X)}$ lesions on variability of song acoustic structure was calculated by the two measures, "within-syllable variability" and "cross-rendition variability" of the fundamental frequency (FF) in song syllables (43). We randomly chose ~50 song motifs recorded on the pre-lesion day and those recorded on the post-lesion day, and extracted only syllables that had clear and flat harmonic structure. For each syllable rendition, a trajectory of FF was obtained in a sound segment of harmonic structure as in a previous study (41). Briefly, spectrograms were calculated using a Gaussian-windowed short-time Fourier transform (σ = 1 ms) sampled at 8 kHz, and a trajectory of the FF (the 1st harmonic frequency) was obtained by calculating the FF in individual time bins. For a subset of syllables that exhibits relatively low signal-to-noise ratios in the 1st harmonic frequency, the 2nd or upper harmonic frequency was used to quantify the FF trajectory. For each syllable, FF trajectories of all renditions were aligned by the onset of the syllables, based on amplitude-threshold crossings, and flat portions ($\geq$ 25 ms) of FF trajectories were used for further analysis. We first removed the modulation of FF trajectories that was consistent across renditions by calculating residual FF trajectories as percent deviation from the mean trajectory across renditions. We then obtained within-syllable variability by calculating the SD of FF within each FF trajectory and averaging it across all renditions. To obtain cross-rendition variability, mean FF in each FF trajectory was calculated, and then the SD of mean FF across all renditions was computed.

***In situ* hybridization**

NTS cDNA fragments used for the synthesis of *in situ* hybridization probes were cloned from a whole-brain cDNA mixture of a male zebra finch. Total RNA was transcribed to cDNA using Superscript Reverse Transcriptase (Invitrogen) with oligo dT primers. The cDNAs were amplified by PCR using oligo DNA primers directed to conserved regions of the coding sequence from the NCBI cDNA database (accession # NM_001245684). PCR products were ligated into the pGEM-T Easy plasmid (Promega). The cloned sequences were searched using NCBI BLAST/BLASTX to compare with homologous genes with other species, and genome loci were identified using BLAT of the UCSC Genome Browser. For fluorescence in situ hybridization (FISH), digoxigenin (DIG)-labeled riboprobes were used. A total of 100–200 ng/glass of the DIG-labeled riboprobe was mixed with the hybridization solution [50% formamide, 10% dextran, 1× Denhardt's solution, 1 mM EDTA (pH 8.0), 33 mM Tris-HCl (pH 8.0), 600 mM NaCl, 0.2 mg/mL yeast tRNA, 80 mM dithiothreitol, and 0.1% N-lauroylsarcosine]. Hybridization was performed at 65 °C for 12–14 h. Washing steps were

performed as follows: 5× SSC solution at 65 °C for 30 min, formamide-I solution (4× SSC, 50% formamide, and 0.005% Tween20) at 65 °C for 40 min, formamide-II solution (2× SSC, 50% formamide, and 0.005% Tween20) at 65 °C for 40 min, 0.1× SSC at 65 °C for 15 min × 2, 0.1× SSC at RT for 15min, NTE buffer at RT for 20 min, and TNT buffer × 3, and blocking buffer [1% DIG blocking solution (Roche) + 1% normal goat serum/1× TNT buffer] at RT for 30 min. DIG-labeled probes were detected with anti-DIG HRP-conjugated antibody (Jackson Laboratory) and a TSA Plus Cy5 system (Perkin Elmer). Signal images were obtained by fluorescence microscopy (EVOS FL; Thermo Fisher Science; BZ-X700; KEYENCE).

The number of $HVC_{(X)}$ neurons was estimated as the average NTS+ cells/mm^2 in both hemispheres of individuals. On the based of the value of NTS+ cells/mm^2, the degree of ablation of $HVC_{(X)}$ neurons in individual birds was calculated as a normalized value (%) with the average of NTS+ cells/mm^2 of control birds.

Adeno-associated virus (AAV) construction

All the viral ITR-flanked genomes used in this study were of the self-complementary (sc) AAV vector type (44). The pscAAV-GFP vector containing a CMV promoter was obtained from Addgene (#32396). AAV plasmids containing Cre and DIO (double-floxed inverted open reading frame)/FLEx (Flip excision) inserts were obtained from Dr. Kenta Kobayashi from the National Institute of Physiological Sciences and subsequently cloned into the pscAAV vector plasmids after amplification of the Cre and DIO/FLEx sequences by primers containing the corresponding restriction enzymes in the target plasmid. To cell-specifically ablate the $HVC_{(X)}$ cells, a combination of diphtheria toxin A (dtA) and constitutively active caspase 3 was used (45-47). Diphtheria toxin was cloned from pAAV-mCherry-FLEx-dtA (Addgene, #58536) by primers with specific enzyme sites and inserted into the previously constructed scAAV-DIO/FLEx. Owing to the restricted carrying capacity of the pscAAV vector, it became necessary to generate a constitutively active caspase 3 (47) by insertional mutagenesis of rAAV-flex-taCasp3-Tevp obtained from Gene Therapy Center Vector Core at the University of North Carolina at Chapel Hill. This insertion consisted of the substitution of valine with glutamic acid at residue 266 of the protein, with subsequent amplification and cloning into an scAAV-DIO/FLEx vector.

AAVs were produced in-house using AAVpro 293T (Takara) cells transfected with a polyethyleneimine (PEI)-condensed recombinant DNA mixture, based on a protocol kindly provided to us by the Gradinaru Lab at Caltech. AAVpro 293T cells were amplified in 10 cm sterile culture dishes under standard cell culture medium [D-MEM, 10% fetal bovine serum

(FBS), 1% penicillin, 1% GlutaMAX] until at least 1.5×10^8 cells could be collected. Cells were then plated onto 15 plates of 15 cm culture dishes at 1.0×10^7 cells per plate in standard cell culture medium. The confluence of the 15cm plates was checked visually, and the medium was changed again to standard 10% FBS medium when the confluence reached 80%, before proceeding to the transfection step. The transfection mix contained a triple plasmid system (pPack2/9 for serotype 2 Rep and serotype 9 Cap genes, pHelper for the adenoviral helper genes and the ITR-flanked viral genome containing plasmid) mixed with 40 kDa PEI in a 1: 3.5 DNA: PEI weight ratio, all dissolved in warm PBS. The cells were maintained in this transfection mix for 24 h, and then the medium was changed to a low serum one (D-MEM, 5% FBS, 1% penicillin, 1% GlutaMAX) to promote protein synthesis instead of cell division. Cells were scraped and collected in buffer (150 mM NaCl, 100 mM Tris-HCl pH 8.0) 3 days after transfection and maintained in a –80 °C freezer until viral purification. The purification procedure was performed as follows. Cells were freeze-thawed between a 37 °C water bath and a –80 °C cooled ethanol bath for at least four cycles and then incubated for 30 min in Benzonase (Merck-Millipore) nuclease after adding of 60 µL of 1 M $MgCl_2$ to the cell solution. At the end of this incubation, the cell solution was centrifuged at 4°C and 7,000 rpm for 1 h at 4°C, and then its supernatant was added as the top layer of a polycarbonate centrifuge column filled with an iodixanol gradient (15%, 25%, 40%, and 54% iodixanol layers). After ultracentrifugation for 6 h at 28,000 rpm, the 40% layer was extracted with a syringe and concentrated in four VivaSpin (Sartorius) cycles. Samples were finally aliquoted and stored in PBSF at –80 °C.

Surgery

Virus injection surgeries were performed on a custom-modified stereotaxic apparatus under 0.6–2.0% isoflurane anesthesia. To locate HVC and Area X, both stereotaxic coordinates from the midsagittal sinus “y point” (0 mm rostral–caudal and 2.0–2.2 mm medial–lateral from the y point for HVC, 7.8 mm rostral–caudal and 1.5 mm medial–lateral from the y point for Area X) and electrophysiological measurements using 1 M NaCl backfilled glass capillaries attached to a recording-capable Nanoject II (Drummond) were used. The location of injection sites for juvenile birds was slightly different (roughly 0.3 mm shallower for Area X and closer to the midsagittal sinus for HVC), and special care was taken to shorten the surgery time as much as possible. The viral solution (virus titer 5.0×10^{12} to 5.1×10^{13} Vg/mL, a total of 1 µL in each Area X, and 800 nL in each HVC) was injected with a pressure Nanojector II. Deafening surgery was performed on the birds by cochlear extirpation after crystallization at phd 104–110 for the adult deaf $HVC_{(X)}$ ablation experiment. The birds were anesthetized with pentobarbital (6.48

mg/mL; 60 µL/10 g body weight) by intraperitoneal injection. After fixing the head in a custom-made stereotaxic apparatus with ear bars, a small window was made through the neck muscle and the skull near the end of the elastic extension of the hyoid bone. A small hole was then made in the cochlear dome. The cochlea was pulled out with a fine hooked wire. The removed cochleae were confirmed by visual inspection under a dissecting microscope. After cochlear removal, the birds recovered on a heat pad before being put back in their cages.

Acknowledgments

We thank Masahiko Kobayashi for initial AAV production and Keiko Sumida for her excellent bird care and breeding. This work was supported by MEXT/JSPS KAKENHI Grant Number JP17H01015 to K.O. and K.W. and by #4903-JP17H06380, JP19H04888, JP17K19629, and JP18H02520 to K.W.

Author contributions

M.S.-V. and K.W. designed the research. M.S.-V., Y. Su., Y. Sh., Y.J., N.A., and K.W. performed the experiments. C.N.A., I.K., H.O., K.K., and K.O. helped with the initial experiments and material sampling. N.T., D.M., and S.K. helped with song analyses. M.S.-V. and K.W. performed the analysis and wrote the paper.

Competing interests

The authors declare that no competing interests exist.

Data availability

The authors confirm that all of the data for the reported findings are included in the article or in supplementary data files. All raw data are available from the authors upon request.

Figures and figure legends

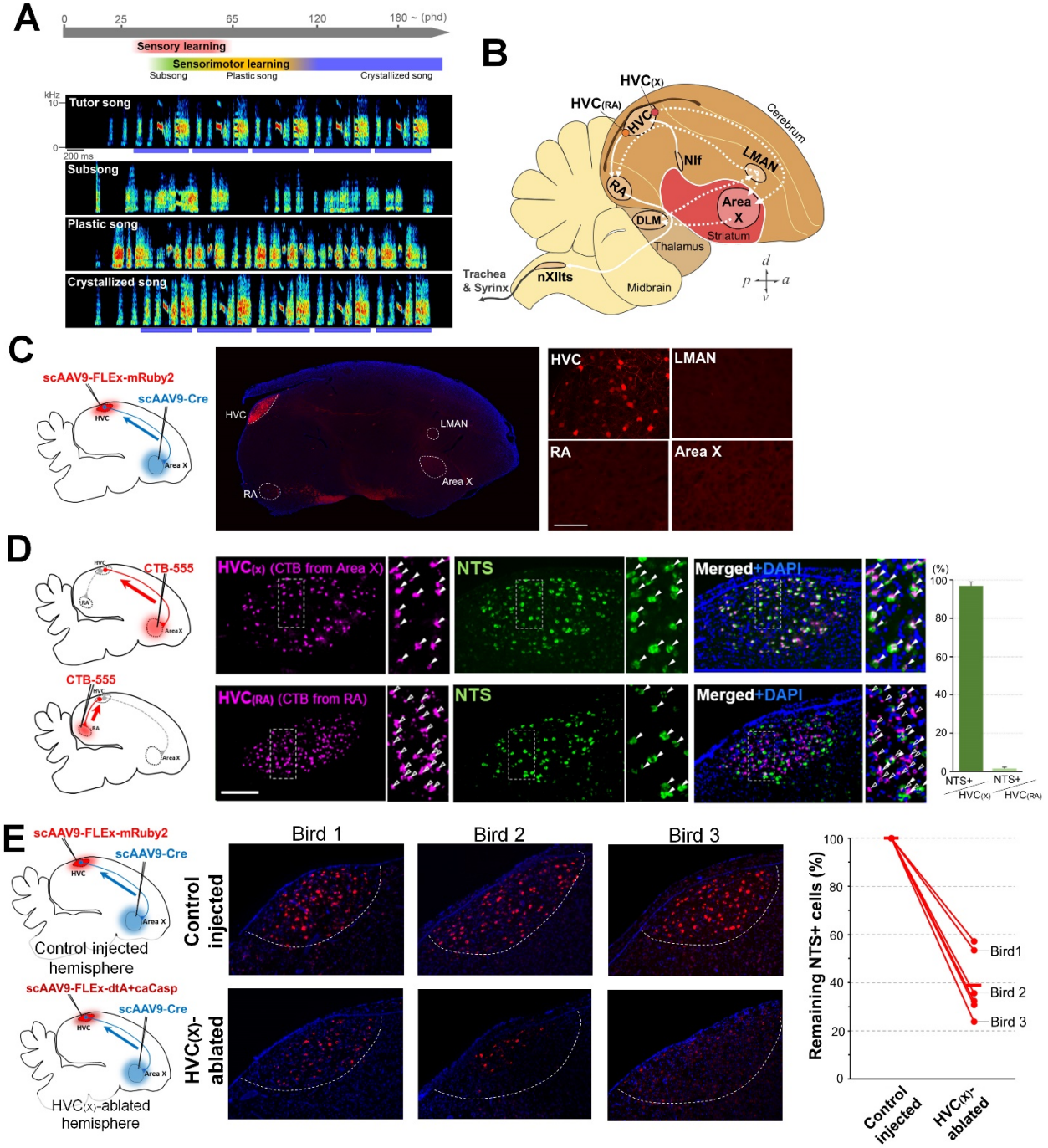


Figure 1. Specific ablation of $HVC_{(X)}$ neurons projecting to the basal ganglia Area X

(A) (Top panel) Time course of song learning in the zebra finch. (Bottom panels) Spectrograms illustrating the progression of song learning. Blue bars represent the motif structure of the crystallized song.

(B) Diagram showing selected song-control regions and connections in the zebra finch brain. The posterior motor pathway and the anterior forebrain pathway (cortical-basal ganglia-thalamic circuit) are represented as solid and dotted white lines, respectively. HVC (used as a proper name); RA, robust nucleus of the arcopallium; Area X, Area X of the striatum; DLM, dorsal lateral nucleus of the medial thalamus; LMAN, lateral magnocellular nucleus of the anterior nidopallium; NIf, interfacial nucleus of the nidopallium; nXIIts, tracheosyringeal part of the hypoglossal nucleus.

(C) (Left diagram) $HVC_{(X)}$ projection neurons were targeted using a combination of retrograding AAV-Cre injected in basal ganglia nucleus Area X and AAV-FLEx-mRuby injected in HVC. (Right panels) Restricted expression of FLEx-inverted mRuby2 fluorescent protein in the $HVC_{(X)}$ cell population. Scale bar = 100μm.

(D) Selective expression of NTS in $HVC_{(X)}$ neurons (green). $HVC_{(X)}$ and $HVC_{(RA)}$ neurons were backfilled with the retrograde tracer CTB-555 from Area X and RA , respectively (magenta). DAPI (blue).

(E) Normalized decreased amount of $HVC_{(X)}$ neurons between control (left) and lesioned HVC. The control hemisphere was injected with scAAV9-Cre in Area X and with scAAV9-FLEx-mRuby2 in HVC. The lesioned hemisphere was injected with scAAV9-Cre in Area X and with a mixture of scAAV9-FLEx-dtA and -caCasp in HVC.

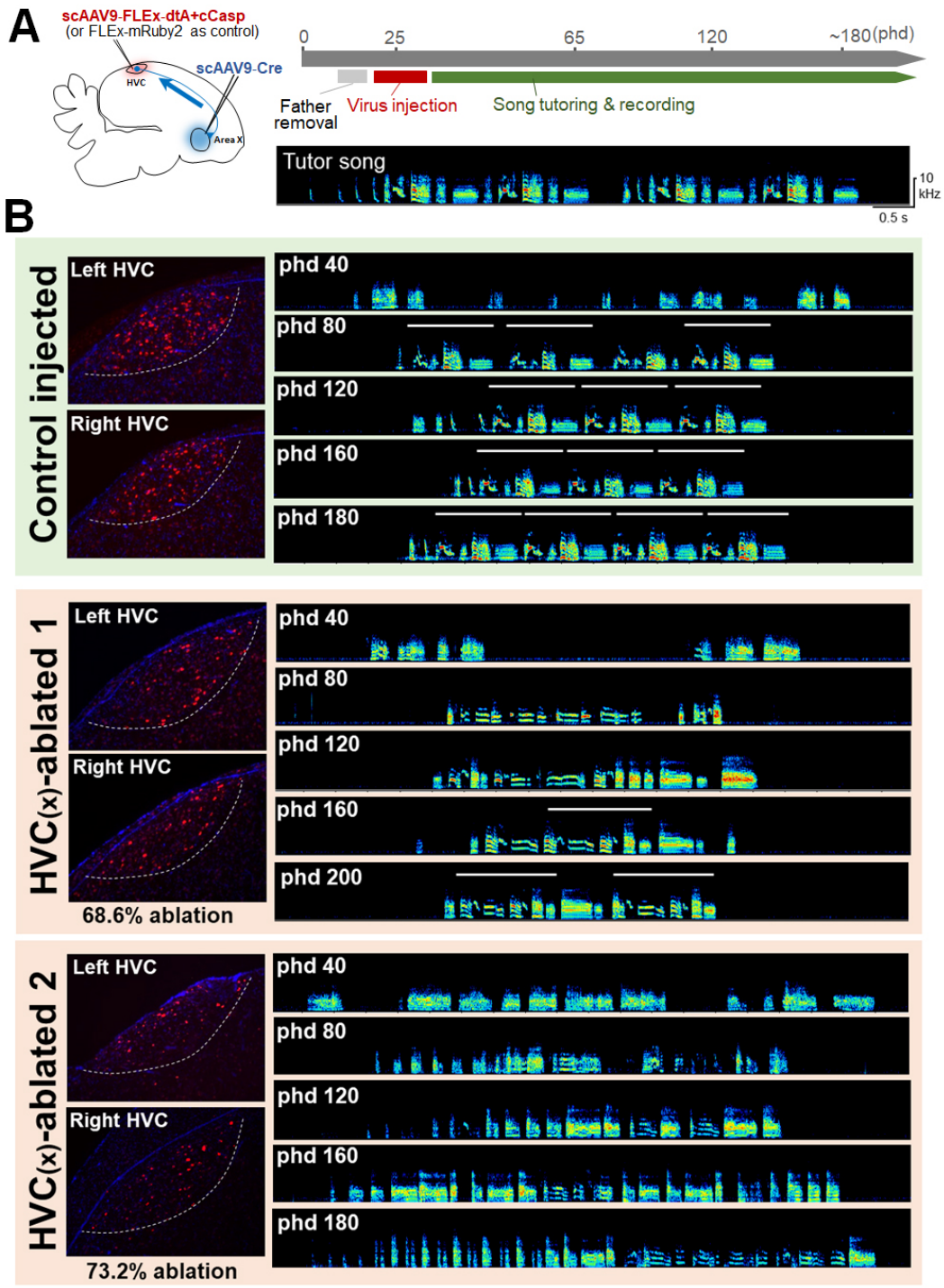


Figure 2. Ablation of $HVC_{(X)}$ neurons in juveniles induces deficits in song learning and development

(A) Experimental timeline for $HVC_{(X)}$ ablation and song tutoring

(B) Examples of song development in a control injected (green-colored background) and two $HVC_{(X)}$-ablated (brown-colored background) birds. $HVC_{(X)}$-ablated birds 1 and 2 had

decreases of 68.6% and 73.2% of $HVC_{(X)}$ neurons, respectively, compared with the average of $HVC_{(X)}$ neurons in the control birds. White lines in the song spectrograms represent the motif structure of songs. The remaining $HVC_{(X)}$ neurons were labeled with NTS (red). DAPI (blue). White dotted lines represent HVC borders.

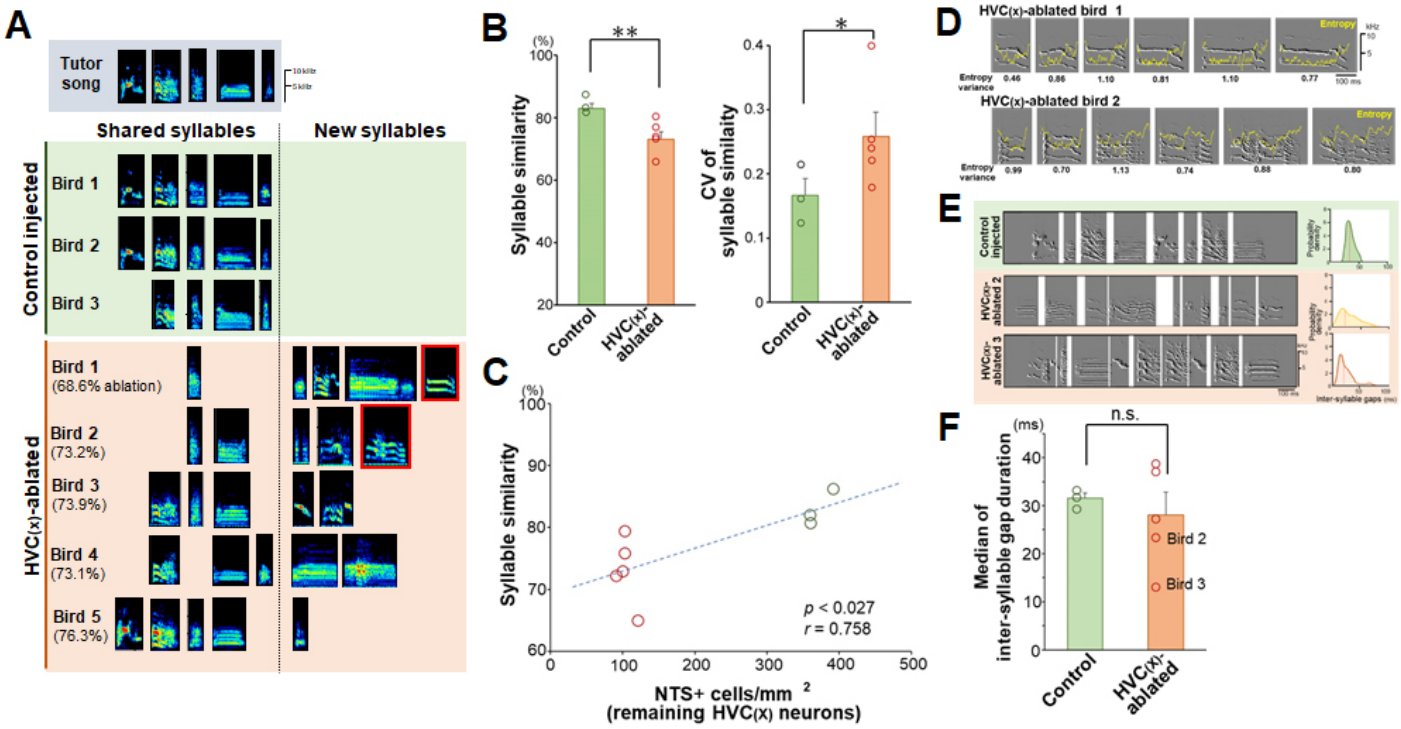


Figure 3. Ablation of HVC$_{(X)}$ neurons in the juvenile stage caused abnormality in syllable acoustics and inter-syllable gap duration in adult songs

(A) Examples of acquired syllables in control (green background) and HVC$_{(X)}$-ablated (brown-colored background) birds. Syllables outlined with red lines were further analyzed in panel (**C**).

(B) Differences between control and HVC$_{(X)}$-ablated birds in the syllable similarity between syllables of each pupil and the tutor song (left) and its CV (right) (n = 3 controls, n = 5 ablated birds; Student's *t* test: **p* < 0.05, ***p* < 0.01). Mean + SEM for bar graphs. (Left) Each point represents the average similarity score of all syllable types for individual birds. (Right) Each point represents the CV of the similarity scores of all syllable types for individual birds.

(C) Correlation between NTS+ cell density in HVC (*i.e.*, degree of residual HVC$_{(X)}$ neurons) and syllable similarity between syllables of each pupil and the tutor song ($p < 0.027$, $r = 0.758$, Pearson's correlation coefficient). Green and red circles represent control and HVC$_{(X)}$-ablated birds, respectively.

(D) High variability in duration and acoustics in syllables in the adult stage (phd 150) for birds whose HVC$_{(X)}$ neurons were ablated in the juvenile stage. Yellow lines represent acoustic entropy, and numerical values show entropy variance.

(E) Examples of abnormal inter-syllable gaps in the adult stage for birds whose HVC$_{(X)}$ neurons were ablated in the juvenile stage. (Left) Variability and shortening of inter-

syllable gaps in $HVC_{(X)}$-ablated birds in the juvenile stage. (Right) Probability density of inter-syllable gaps from each bird (n = 100 gaps). The red dotted lines indicate median values.

(F) Median of inter-syllable gap duration between control and $HVC_{(X)}$-ablated birds (100 inter-syllable gaps/bird). Each dot represents an individual bird's value. Bird ID numbers are consistent between panels **(A)**, **(C)**, **(D), (E), and (F)**.

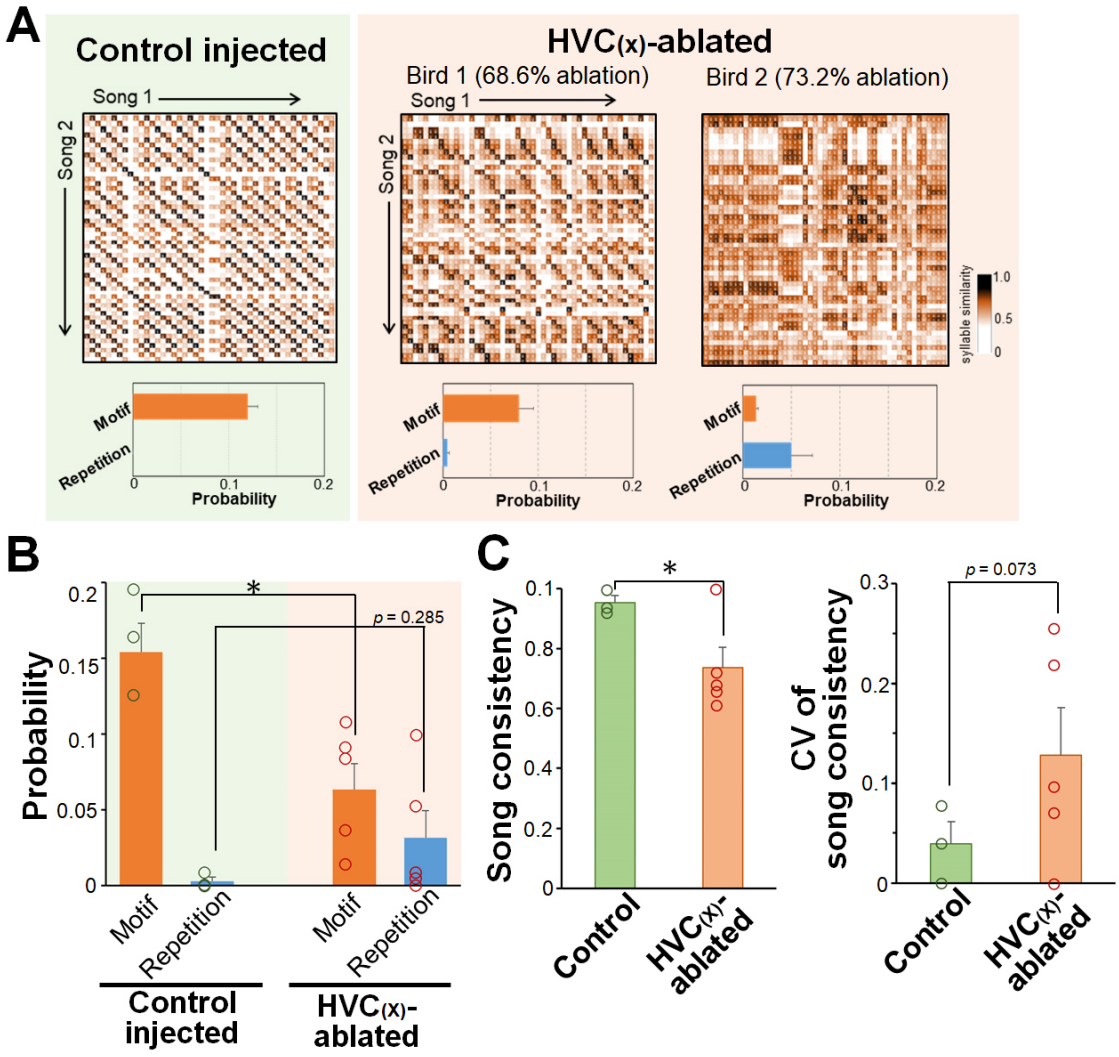


Figure 4. Altered sequential properties of the adult songs of HVC$_{(X)}$-ablated birds in the juvenile stage

(A) (Top) Representative syllable similarity matrices (SSMs) for adult songs (phd 150) in control (green background) and two HVC$_{(X)}$-ablated (brown background) birds. (Bottom) Probabilities of motif and repetition indices for each bird.

(B) Probabilities of motif and repetition indices in the adult stage (phd 150) in control and HVC$_{(X)}$-ablated birds (n = 3 controls, n = 5 ablated birds; Student's *t* test: **p* < 0.05). Dots indicate individual bird's values.

(C) Song sequence consistency and its CV at phd 150 in control and HVC$_{(X)}$-ablated birds (n = 3 controls, n = 5 ablated birds; Student's *t* test: **p* < 0.05). Mean + SEM for all graphs. Dots indicate individual bird's values.

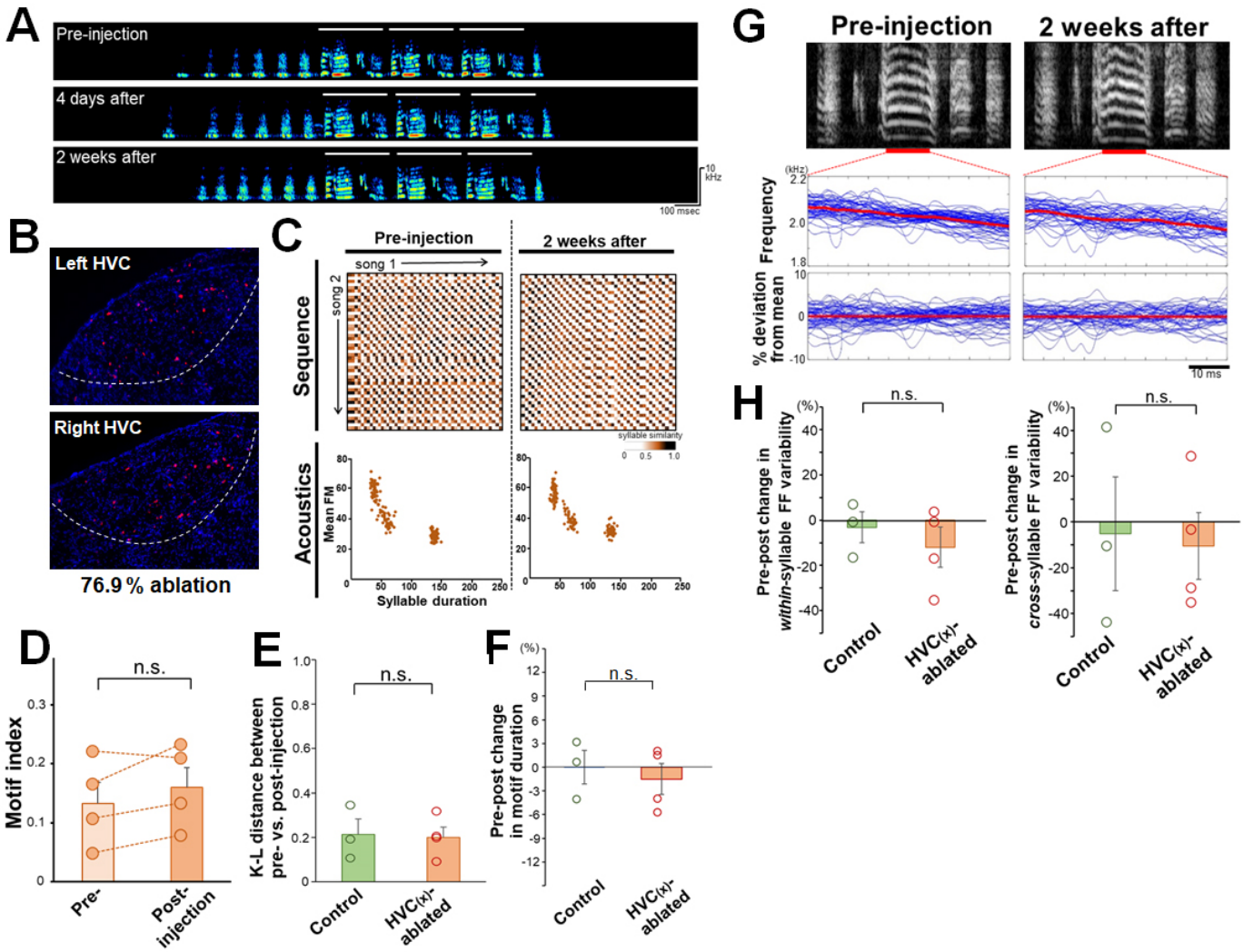


Figure 5. Non-obvious change in song structure by ablation of HVC$_{(X)}$ neurons in adults

(A) Representative spectrogram of birds that were ablated in HVC$_{(X)}$ neurons in adults. White bars represent the motif structure of songs.

(B) Example of the extent of HVC$_{(X)}$ ablation in an ablated adult (with 76.9% ablation) as shown in **(A)** and **(C)**. NTS (red) and DAPI (blue).

(C) Syllable sequence and acoustic stability before and after ablation of HVC$_{(X)}$ neurons. Sequential patterns are shown as SSMs and acoustics as a scatter density plot of syllable duration versus mean FM (n = 150 syllables).

(D) No effect of HVC$_{(X)}$ ablation on the song motif index of adult zebra finches. Each point corresponds to an individual bird.

(E) No effect of HVC$_{(X)}$ ablation on syllable acoustics measured by the K-L distance of syllable scatter density plots (duration versus mean FM) between pre- and post-injection time points

(n = 3 controls, n = 4 ablated birds; Student's *t* test: n.s.,$p > 0.05$).

(F) Pre–post change in motif duration between control and $HVC_{(X)}$-ablated birds (Student's *t* test: n.s., $p > 0.05$).

(G) Example of an FF trajectory of a syllable in pre-injection (top left) and 2 weeks post-injection (top right) of songs from an $HVC_{(X)}$-ablated adult, expressed as raw frequency traces (middle) and percent deviation from the within-rendition mean (bottom). Blue and red lines indicate each rendition and the mean across renditions, respectively.

(H) Pre–post changes in within- and cross-rendition syllable variability in FF between control and $HVC_{(X)}$-ablated birds (Student's *t* test: n.s.,$p > 0.05$). Mean +SEM for all graphs.

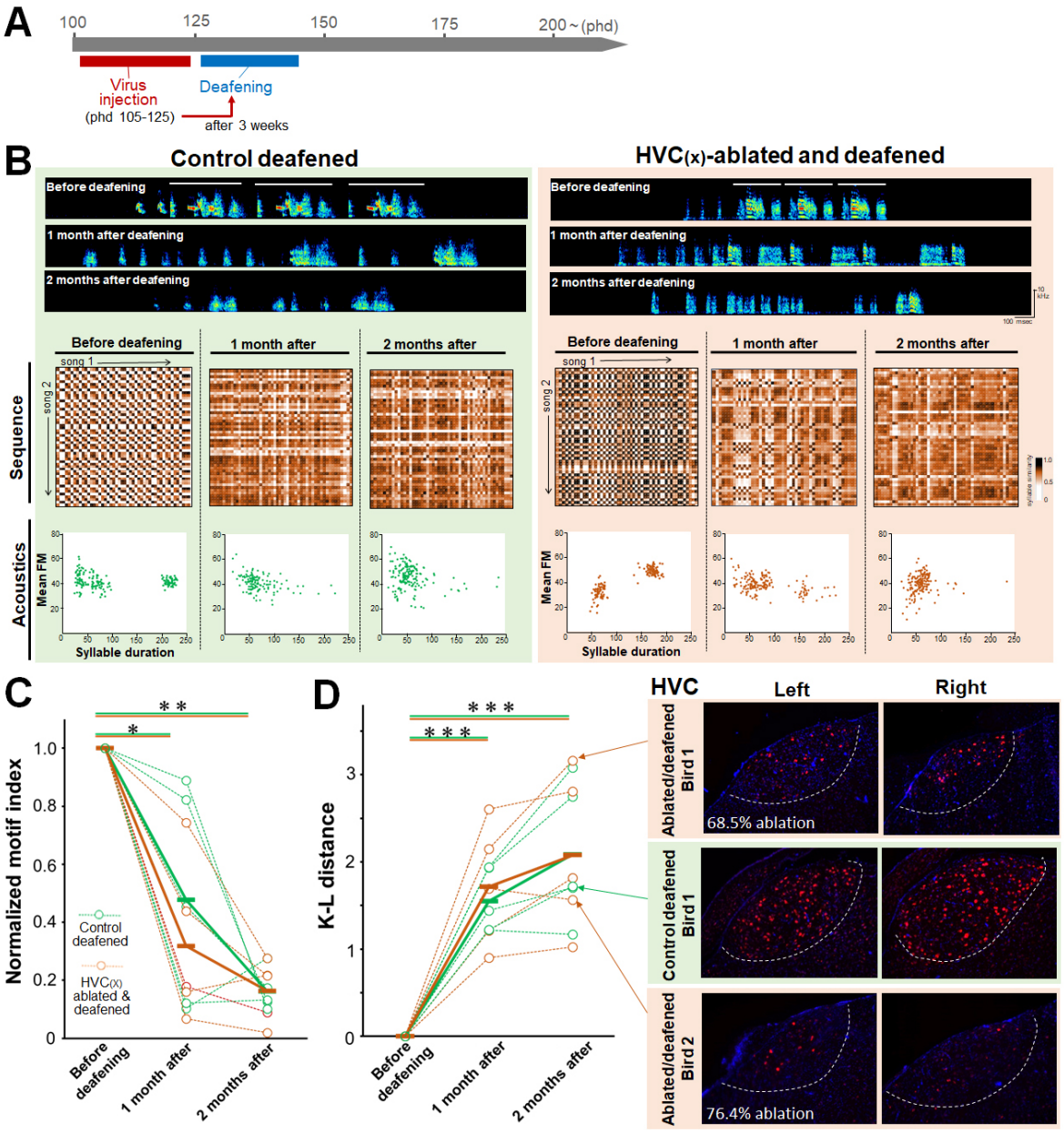


Figure 6. Ablation of HVC$_{(X)}$ neurons does not alter song degradation after deafening in adult zebra finches

(A) Timeline of HVC$_{(X)}$ ablation and deafening in the adult stage.

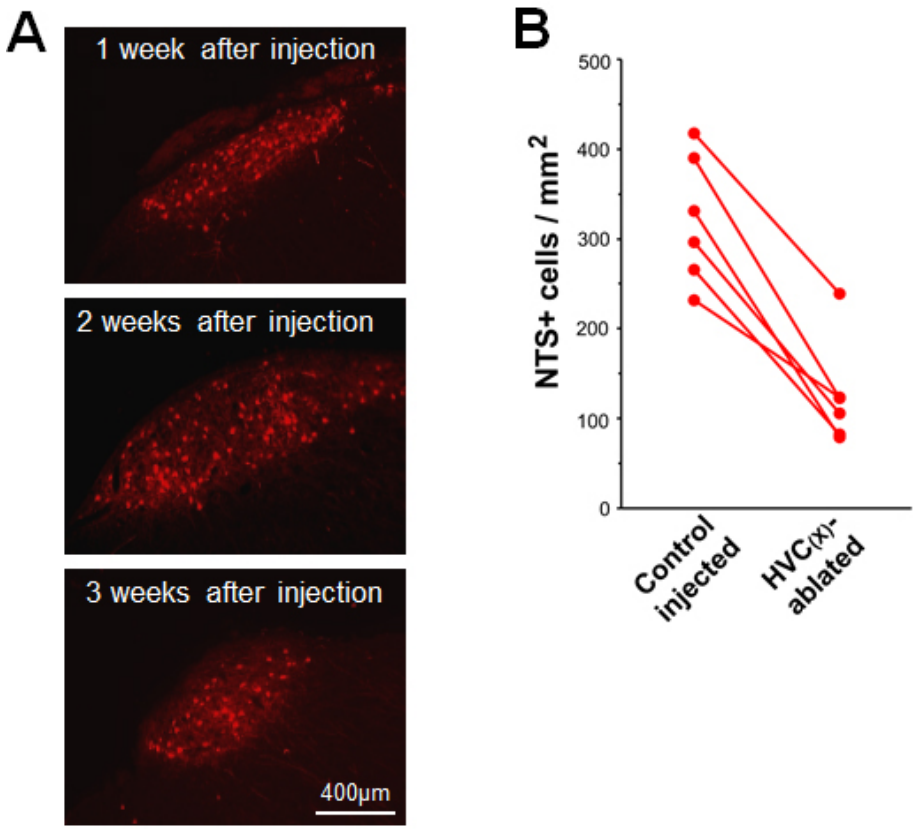
(B) Deafening-induced degradation of the syllable sequence and acoustics in a control (green background) and HVC$_{(X)}$-ablated (brown background) adult birds.

(C) Similar rates of deafening-induced degradation of song motif structure between control and HVC$_{(X)}$-ablated adult birds (n = 5 for each group; paired *t* test: **p* < 0.05, ***p* < 0.01). Green and brown lines represent control and HVC$_{(X)}$-ablated birds, respectively. Dotted- and solid

lines represent individual and average values, respectively.

(D) (Left) Similar rates of acoustic degradation after deafening between control and $HVC_{(X)}$-ablated birds, as calculated by the K–L distance (n = 5 for each group; paired *t* test: ***$p < 0.001$). (Right) Remaining $HVC_{(X)}$ neurons in three representative birds (a control and two $HVC_{(X)}$-ablated birds), visualized by NTS (red) and DAPI (blue). White dotted lines represent the border of HVC.

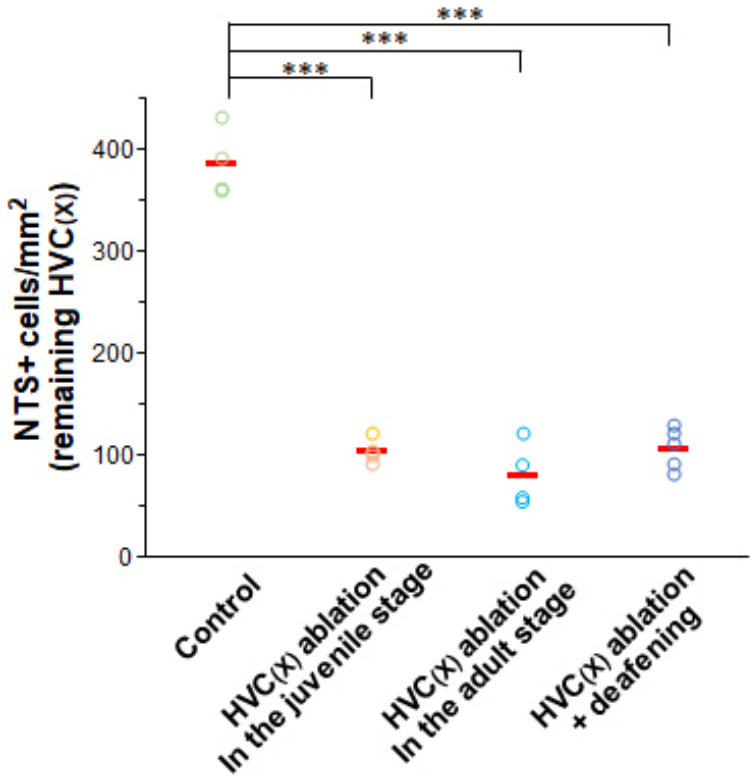
Supplementary figure legends



Supplementary figure 1. Retrograde transported Cre-dependent FLEx inversion timing after scAAV9 injection and $HVC_{(X)}$ neuron ablation

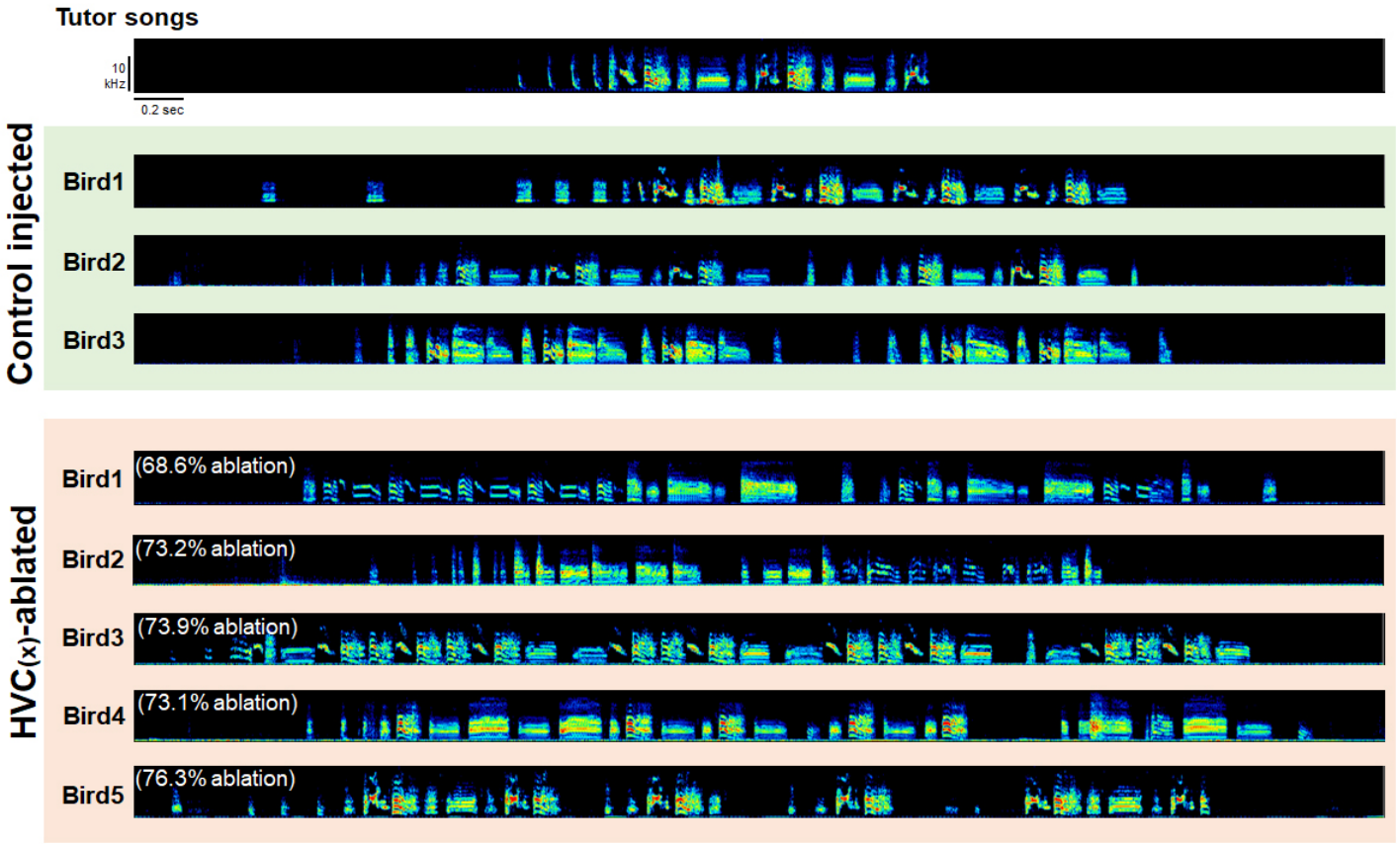
(A) Restricted expression of FLEx-inverted mRuby2 fluorescent protein in $HVC_{(X)}$ cell populations at 1, 2, and 3 weeks after virus injection.

(B) Comparison of $HVC_{(X)}$ neuron density between control and lesioned HVC. The control hemisphere was injected with scAAV9-Cre in Area X and with scAAV9-FLEx-mRuby2 in HVC. The lesioned hemisphere was injected with scAAV9-Cre in Area X and with a mixture of scAAV9-FLEx-dtA and scAAV9-FLEx-caCasp in HVC.



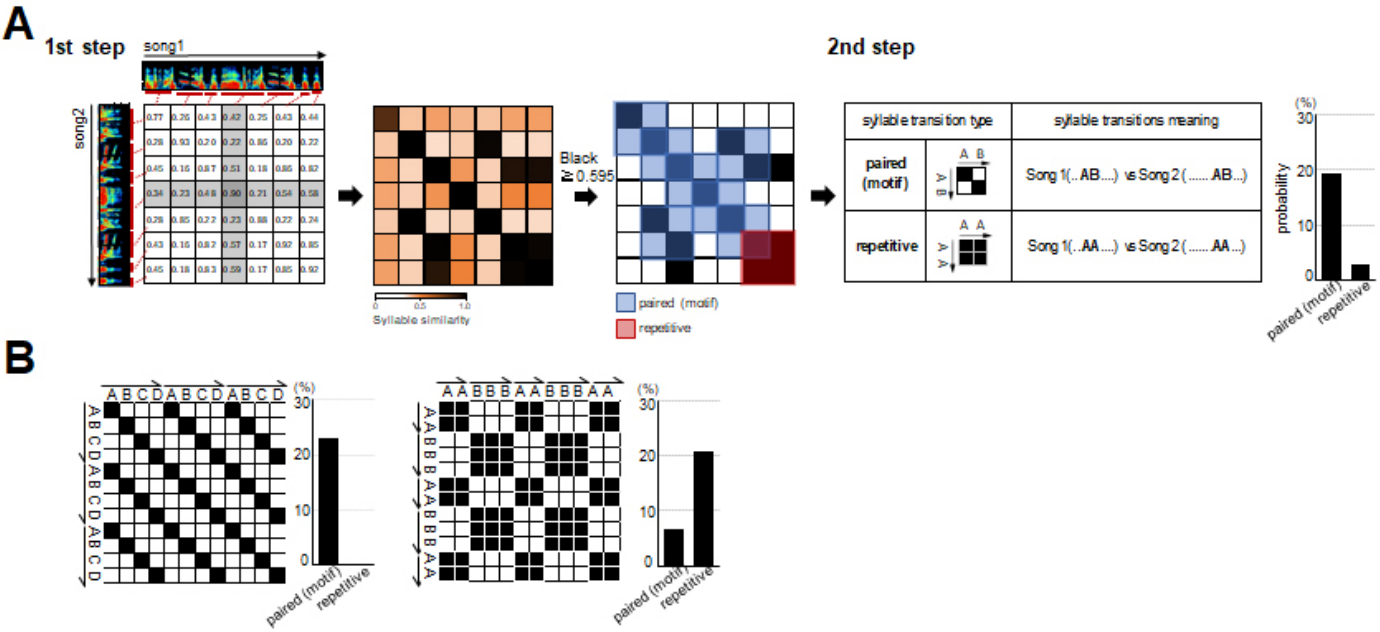
Supplementary figure 2. Comparison of remaining $HVC_{(X)}$ neurons among control, ablation in the juvenile stage, adult stage, and following deafening conditions

Each dot corresponds to the average density of NTS+ cells ($HVC_{(X)}$ neurons) in one bird. Red horizontal bars represent the mean values for each group (Tukey HSD, ***p <1e−07).



Supplementary figure 3. Examples of acquired songs at phd 180 from control and HVC(x)-ablated birds in the juvenile stage

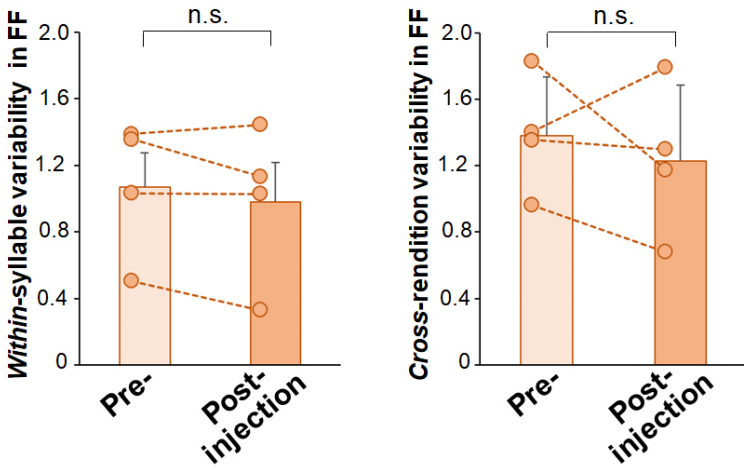
Bird numbers are consistent in Figures 2–4.



Supplementary figure 4. Syllable similarity matrix (SSM) method for the detection of syllable transition patterns

(A) The SSM method consists of two steps. In the first step, a correlation matrix including syllable similarity scores was prepared by the round-robin comparison of all the syllable comparisons in two songs, maintaing the sequential order of the syllables in the songs. These similarity scores in the matrix were binarized at a threshold at 0.595. In the second step, the occurrence rate of two patterns of binarized "2 row × 2 column" cells in the SSM was calculated as a percentage of the paired (motif) and repetitive-syllable transition types (see Materials and methods).

(B) Test examples of the SSM method using artificial song models mimicking the songs with motif and repetitive sequences.



Supplementary figure 5. Within- and cross-rendition syllable variability in fundamental frequency (FF) between pre- and post-2 weeks injection in $HVC_{(X)}$-ablated birds (n = 4; paired *t* test: n.s., $p > 0.05$)

References

1. O. Hikosaka, K. Nakamura, K. Sakai, H. Nakahara, Central mechanisms of motor skill learning. *Current opinion in neurobiology* **12**, 217-222 (2002).
2. J. Tanji, Sequential Organization of Multiple Movements: Involvement of Cortical Motor Areas. *Annual Review of Neuroscience* **24**, 631-651 (2001).
3. M. S. Fee, J. H. Goldberg, A hypothesis for basal ganglia-dependent reinforcement learning in the songbird. *Neuroscience* **198**, 152-170 (2011).
4. T. Tesileanu, B. Olveczky, V. Balasubramanian, Rules and mechanisms for efficient two-stage learning in neural circuits. *eLife* **6** (2017).
5. N. Fujii, A. M. Graybiel, Representation of Action Sequence Boundaries by Macaque Prefrontal Cortical Neurons. *Science* **301**, 1246-1249 (2003).
6. X. Jin, R. M. Costa, Shaping action sequences in basal ganglia circuits. *Current opinion in neurobiology* **33**, 188-196 (2015).
7. T. D. Barnes, Y. Kubota, D. Hu, D. Z. Jin, A. M. Graybiel, Activity of striatal neurons reflects dynamic encoding and recoding of procedural memories. *Nature* **437**, 1158-1161 (2005).
8. A. A. Kozhevnikov, M. S. Fee, Singing-related activity of identified HVC neurons in the zebra finch. *Journal of neurophysiology* **97**, 4271-4283 (2007).
9. H. Fujimoto, T. Hasegawa, D. Watanabe, Neural Coding of Syntactic Structure in Learned Vocalizations in the Songbird. *Journal of Neuroscience* **31**, 10023-10033 (2011).
10. Q. Li *et al.*, Refinement of learned skilled movement representation in motor cortex deep output layer. *Nat Commun* **8**, 15834 (2017).
11. X. Jin, R. M. Costa, Start/stop signals emerge in nigrostriatal circuits during sequence learning. *Nature* **466**, 457-462 (2010).
12. E. D. Stefanova, V. S. Kostic, L. Ziropadja, M. Markovic, G. G. Ocic, Visuomotor skill learning on serial reaction time task in patients with early Parkinson's disease. *Movement disorders : official journal of the Movement Disorder Society* **15**, 1095-1103 (2000).
13. D. B. Willingham, W. J. Koroshetz, Evidence for dissociable motor skills in Huntington's disease patients. *Psychobiology* **21**, 173-182 (1993).
14. P. Marler, Birdsong and speech development: Could there be parallels? There may be basic rules governing vocal learning to which many species conform, including man. *American scientist* **58**, 669-673 (1970).

15. M. Konishi, The role of auditory feedback in the control of vocalization in the white-crowned saprrow. *Zeitschrift fur Tierpsychologie* **22**, 770-783 (1965).
16. M. S. Brainard, A. J. Doupe, Auditory feedback in learning and maintenance of vocal behaviour. *Nature reviews. Neuroscience* **1**, 31-40 (2000).
17. A. J. Doupe, P. K. Kuhl, Birdsong and human speech: common themes and mechanisms. *Annu Rev Neurosci* **22**, 567-631 (1999).
18. E. D. Jarvis, Learned birdsong and the neurobiology of human language. *Annals of the New York Academy of Sciences* **1016**, 749-777 (2004).
19. A. C. Yu, D. Margoliash, Temporal hierarchical control of singing in birds. *Science* **273**, 1871-1875 (1996).
20. R. H. Hahnloser, A. A. Kozhevnikov, M. S. Fee, An ultra-sparse code underlies the generation of neural sequences in a songbird. *Nature* **419**, 65-70 (2002).
21. M. S. Fee, A. A. Kozhevnikov, R. H. Hahnloser, Neural mechanisms of vocal sequence generation in the songbird. *Annals of the New York Academy of Sciences* **1016**, 153-170 (2004).
22. S. W. Bottjer, E. A. Miesner, A. P. Arnold, Forebrain lesions disrupt development but not maintenance of song in passerine birds. *Science* **224**, 901-903 (1984).
23. C. Scharff, F. Nottebohm, A comparative study of the behavioral deficits following lesions of various parts of the zebra finch song system: implications for vocal learning. *The Journal of neuroscience : the official journal of the Society for Neuroscience* **11**, 2896-2913 (1991).
24. F. Sohrabji, E. J. Nordeen, K. W. Nordeen, Selective impairment of song learning following lesions of a forebrain nucleus in the juvenile zebra finch. *Behav Neural Biol.* **53**, 51-63 (1990).
25. A. R. Pfenning *et al.*, Convergent transcriptional specializations in the brains of humans and song-learning birds. *Science* **346**, 1256846 (2014).
26. F. Nottebohm, T. M. Stokes, C. M. Leonard, Central control of song in the canary, Serinus canarius. *The Journal of comparative neurology* **165**, 457-486 (1976).
27. D. Aronov, A. S. Andalman, M. S. Fee, A specialized forebrain circuit for vocal babbling in the juvenile songbird. *Science* **320**, 630-634 (2008).
28. T. F. Roberts, S. M. Gobes, M. Murugan, B. P. Olveczky, R. Mooney, Motor circuits are required to encode a sensory model for imitative learning. *Nature neuroscience* **15**, 1454-1459 (2012).
29. D. S. Vicario, F. Nottebohm, Organization of the zebra finch song control system: I. Representation of syringeal muscles in the hypoglossal nucleus. *The Journal of comparative neurology* **271**, 346-354 (1988).

30. J. M. Wild, The avian nucleus retroambigualis: a nucleus for breathing, singing and calling. *Brain research* **606**, 319-324 (1993).
31. P. Dutar, H. M. Vu, D. J. Perkel, Multiple cell types distinguished by physiological, pharmacological, and anatomic properties in nucleus HVc of the adult zebra finch. *Journal of neurophysiology* **80**, 1828-1838 (1998).
32. M. Kubota, I. Taniguchi, Electrophysiological characteristics of classes of neuron in the HVc of the zebra finch. *Journal of neurophysiology* **80**, 914-923 (1998).
33. J. F. Prather, S. Peters, S. Nowicki, R. Mooney, Precise auditory-vocal mirroring in neurons for learned vocal communication. *Nature* **451**, 305-310 (2008).
34. G. F. Lynch, T. S. Okubo, A. Hanuschkin, R. H. Hahnloser, M. S. Fee, Rhythmic Continuous-Time Coding in the Songbird Analog of Vocal Motor Cortex. *Neuron* **90**, 877-892 (2016).
35. M. A. Long, D. Z. Jin, M. S. Fee, Support for a synaptic chain model of neuronal sequence generation. *Nature* **468**, 394-399 (2010).
36. M. A. Picardo *et al.*, Population-Level Representation of a Temporal Sequence Underlying Song Production in the Zebra Finch. *Neuron* **90**, 866-876 (2016).
37. K. Hamaguchi, K. A. Tschida, I. Yoon, B. R. Donald, R. Mooney, Auditory synapses to song premotor neurons are gated off during vocalization in zebra finches. *eLife* **3**, e01833 (2014).
38. D. Vallentin, M. A. Long, Motor Origin of Precise Synaptic Inputs onto Forebrain Neurons Driving a Skilled Behavior. *Journal of Neuroscience* **35**, 299-307 (2015).
39. C. Scharff, J. R. Kirn, M. Grossman, J. D. Macklis, F. Nottebohm, Targeted neuronal death affects neuronal replacement and vocal behavior in adult songbirds. *Neuron* **25**, 481-492 (2000).
40. A. S. Andalman, M. S. Fee, A basal ganglia-forebrain circuit in the songbird biases motor output to avoid vocal errors. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 12518-12523 (2009).
41. J. D. Charlesworth, T. L. Warren, M. S. Brainard, Covert skill learning in a cortical-basal ganglia circuit. *Nature* **486**, 251-255 (2012).
42. E. Hisey, M. G. Kearney, R. Mooney, A common neural circuit mechanism for internally guided and externally reinforced forms of motor learning. *Nature neuroscience* **21**, 589-597 (2018).
43. S. Kojima, M. H. Kao, A. J. Doupe, M. S. Brainard, The Avian Basal Ganglia Are a Source of Rapid Behavioral Variation That Enables Vocal Motor Exploration. *The Journal of neuroscience : the official journal of the Society for Neuroscience* **38**, 9635-9647 (2018).
44. D. M. McCarty, P. E. Monahan, R. J. Samulski, Self-complementary recombinant adeno-

associated virus (scAAV) vectors promote efficient transduction independently of DNA synthesis. *Gene therapy* **8**, 1248-1254 (2001).

45. D. M. Lin, V. J. Auld, C. S. Goodman, Targeted neuronal cell ablation in the Drosophila embryo: pathfinding by follower growth cones in the absence of pioneers. *Neuron* **14**, 707-715 (1995).
46. E. Foster *et al.*, Targeted ablation, silencing, and activation establish glycinergic dorsal horn neurons as key components of a spinal gate for pain and itch. *Neuron* **85**, 1289-1304 (2015).
47. J. Walters *et al.*, A constitutively active and uninhibitable caspase-3 zymogen efficiently induces apoptosis. *Biochem J* **424**, 335-345 (2009).
48. A. Kageyama, I. Kusano, T. Tamura, T. Oda, T. Muramatsu, Comparison of the apoptosis-inducing abilities of various protein synthesis inhibitors in U937 cells. *Bioscience, biotechnology, and biochemistry* **66**, 835-839 (2002).
49. N. Komatsu, T. Oda, T. Muramatsu, Involvement of both caspase-like proteases and serine proteases in apoptotic cell death induced by ricin, modeccin, diphtheria toxin, and pseudomonas toxin. *The Journal of Biochemistry* **124**, 1038-1044 (1998).
50. T. F. Roberts *et al.*, Identification of a motor-to-auditory pathway important for vocal learning. *Nature neuroscience* 10.1038/nn.4563 (2017).
51. R. Imai *et al.*, A quantitative method for analyzing species-specific vocal sequence pattern and its developmental dynamics. *Journal of neuroscience methods* 10.1016/j.jneumeth.2016.06.023 (2016).
52. S. Kojima, M. H. Kao, A. J. Doupe, Task-related "cortical" bursting depends critically on basal ganglia input and is linked to vocal plasticity. *Proceedings of the National Academy of Sciences of the United States of America* 10.1073/pnas.1216308110 (2013).
53. F. Ali *et al.*, The Basal Ganglia Is Necessary for Learning Spectral, but Not Temporal, Features of Birdsong. *Neuron* 10.1016/j.neuron.2013.07.049 (2013).
54. W. Wu, J. A. Thompson, R. Bertram, F. Johnson, A statistical method for quantifying songbird phonology and syntax. *Journal of neuroscience methods* **174**, 147-154 (2008).
55. E. Ohgushi, C. Mori, K. Wada, Diurnal oscillation of vocal development associated with clustered singing by juvenile songbirds. *The Journal of experimental biology* **218**, 2260-2268 (2015).
56. L. Kubikova *et al.*, Basal ganglia function, stuttering, sequencing, and repair in adult songbirds. *Scientific reports* **4**, 6590 (2014).
57. M. H. Kao, A. J. Doupe, M. S. Brainard, Contributions of an avian basal ganglia-forebrain circuit to real-time modulation of song. *Nature* **433**, 638-643 (2005).

58. B. P. Ölveczky, A. S. Andalman, M. S. Fee, Vocal Experimentation in the Juvenile Songbird Requires a Basal Ganglia Circuit. *PLoS Biol.* **3**, e153 (2005).

59. M. S. Brainard, A. J. Doupe, Interruption of a basal ganglia-forebrain circuit prevents plasticity of learned vocalizations. *Nature* **404**, 762-766 (2000).

60. M. S. Fee, C. Scharff, The songbird as a model for the generation and learning of complex sequential. *ILAR journal / National Research Council, Institute of Laboratory Animal Resources* **51**, 362-377 (2010).

61. L. Xiao *et al.*, A Basal Ganglia Circuit Sufficient to Guide Birdsong Learning. *Neuron* 10.1016/j.neuron.2018.02.020 (2018).

62. V. Gadagkar *et al.*, Dopamine neurons encode performance error in singing birds. *Science* **354**, 1278-1282 (2016).

63. M. Luo, L. Ding, D. J. Perkel, An avian basal ganglia pathway essential for vocal learning forms a closed topographic loop. *The Journal of neuroscience : the official journal of the Society for Neuroscience* **21**, 6836-6845 (2001).

64. N. I. Denisenko-Nehrbass, E. Jarvis, C. Scharff, F. Nottebohm, C. V. Mello, Site-specific retinoic acid production in the brain of adult songbirds. *Neuron* **27**, 359-370 (2000).

65. T. C. Roeske, C. Scharff, C. R. Olson, A. Nshdejan, C. V. Mello, Long-distance retinoid signaling in the zebra finch brain. *PloS one* **9**, e111722 (2014).

66. B. B. Scott, T. Gardner, N. Ji, M. S. Fee, C. Lois, Wandering neuronal migration in the postnatal vertebrate forebrain. *The Journal of neuroscience : the official journal of the Society for Neuroscience* **32**, 1436-1446 (2012).

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

68. S. Krishnan, K. E. Watkins, D. V. Bishop, Neurobiological Basis of Language Learning Difficulties. *Trends in cognitive sciences* 10.1016/j.tics.2016.06.012 (2016).

69. P. Deriziotis, S. E. Fisher, Speech and Language: Translating the Genome. *Trends in genetics : TIG* 10.1016/j.tig.2017.07.002 (2017).

70. O. Tchernichovski, F. Nottebohm, C. E. Ho, B. Pesaran, P. P. Mitra, A procedure for an automated measurement of song similarity. *Animal behaviour* **59**, 1167-1176 (2000).