Title
Network-based Insights to Learned Vocalization

Author
Burkett, Zachary Daniel

Publication Date
2017

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Network-based Insights to Learned Vocalization

A dissertation submitted in partial satisfaction of the requirements for the degree Doctor of Philosophy in Molecular, Cellular, and Integrative Physiology

by

Zachary Daniel Burkett

2017

ABSTRACT OF THE DISSERTATION

Network-based Insights to Learned Vocalization

by

Zachary Daniel Burkett

Doctor of Philosophy in Molecular, Cellular, & Integrative Physiology

University of California, Los Angeles, 2017

Professor Xinshu Xiao, Co-Chair

Professor Stephanie Ann White, Co-Chair

Being social animals, we as humans can fully appreciate how disorders affecting speech, language, and our ability to communicate severely degrade quality of life. Language is a complex behavior that exists only in humans, making the study of the underlying molecular components challenging. The vocal learning subcomponent of language, however, is shared by a handful of animal taxa, including, among birds, the zebra finch (*Taeniopygia guttata*) songbird species. Beyond this, humans and zebra finches share remarkable parallels with respect to vocalization: our neural circuitry, developmental timelines, and a reliance upon the FoxP2 transcription factor are similar, making them the preeminent model system in which to investigate the molecular basis for learned vocalization with hope of drawing meaningful parallels with humans.

In this dissertation, I describe the process by which I begin untangling the complicated molecular basis for vocal learning which, in humans, is exemplified by speech and language. Mutations in the FOXP2 gene cause a speech and language disorder. Like humans, zebra finches require FoxP2 to properly learn their vocalizations. FoxP2 is down-regulated concurrent with singing behavior in a basal ganglia brain region, Area X, when zebra finches practice their songs. I overexpressed two major isoforms of FoxP2 in the zebra finch brain at a developmentally significant time point wherein the bird is undergoing the song learning process, breaking the link between FoxP2 and singing behavior. In doing this, I discovered unique roles for each isoform: the full-length version contributes strongly to overall vocal learning and variability while the truncated version exerts a strong effect on variability but does not affect learning.

To uncover the molecular basis for these learning and variability phenotypes, I used weighted gene coexpression network analysis (WGCNA) on RNA transcripts from Area X and the outlying non-song ventral striatopallidum (VSP) of animals overexpressing the FoxP2 isoforms or the reporter gene GFP as a control. In the Area X network, modules correlated to singing, learning, and variability. Notably, a large, densely interconnected module positively correlated to learning was discovered. Through comparative network analysis with the non-song juvenile VSP and adult Area X, I discovered the learning related module is present in juvenile VSP but not adult Area X. Further, singing related modules were preserved between juvenile and adult Area X but not between juvenile Area X and VSP. Together, these results indicate a special confluence of singing and learning-related coexpression in juvenile Area X. I then use this information as a model wherein the building blocks of a complex behavior are discrete coexpression patterns. In this case,

the combination of "learning" and "singing" coexpression that occur in juvenile Area X drives the learning behavior.

The correlation of gene expression to behavior is only useful when both gene expression and behavior are accurately quantified. With the advent of RNA-seq, the quantification of gene expression reached a pinnacle. To quantify behavior, I applied principles of WGCNA to sound spectral data, creating the "Vocal Inventory Clustering Engine" (VoICE), which generates clusters of bird vocalizations in an unbiased fashion, a task not possible with existing song analysis software. As part of multiple collaborations, I applied this methodology to the ultrasonic vocalizations of mice, creating, for the first time, a software solution for grouping the variable vocal repertoires of rodents into discrete vocalization "types" in an unbiased and semi-automated fashion. To demonstrate VoICE's utility, I replicated prior work where *Cntnap2* deletion diminished the amount of calling behavior in mouse pups then used VoICE to describe how the knockout makes the vocal repertoire more simple. By using the same network-based principles to group and describe both avian and rodent vocalizations, VoICE allows for a cross-species approach to be taken in determining the relationship between genes and behavior.

The dissertation of Zachary Daniel Burkett is approved.

Giovanni Coppola

Xia Yang

Xinshu Xiao, Committee Co-Chair

Stephanie Ann White, Committee Co-Chair

University of California, Los Angeles

2017

DEDICATION

I have learned that the importance comes not from “how much” but instead “with whom.” I have been fortunate to have a supportive family, a loving girlfriend, and an incredible group of scientists and friends that make up the network of my life. To them I dedicate this work.

Table of Contents

List of Tables

List of Figures

Acknowledgements

I am deeply grateful to Dr. Stephanie White for her support, mentorship, and guidance through both of my graduate degrees. My dissertation committee, Drs. Xinshu Xiao, Xia Yang, and Giovanni Coppola were immensely helpful in the design of the gene network study and interpretation of the results. Their ideas, feedback, and encouragement following the assembly of the dissertation committee were and continue to be greatly appreciated. The contributions of Dr. Austin Hilliard to my education in gene coexpression network analysis were indispensable, noteworthy, and greatly appreciated. The feedback, support, and general good nature of my compatriots in the White lab throughout my graduate career were essential to my development as a scientist, experimenter, and person. I will remain thankful that I had the opportunity to work with and learn from them. These people include but are not limited to: Dr. Julie Miller, Dr. Nancy Day, Dr. Austin Hilliard, Dr. Michael Condro, Dr. Jonathan Heston, Dr. Carmen Panaitof, Elizabeth Fraley, George Hafzalla, Caitlin Aamodt, Madza Virgens, Todd Kimball, Jennifer Morales, Lily Sung, Aneesa Yousefi, Donald Noble, Ryan Dosumu-Johnson, and Diana Sanchez-Gomez.

Chapter 1 is an introduction to the field I entered when beginning the project that is summarized in the second chapter of this dissertation.

Chapter 2 is a version of a manuscript currently in preparation: Burkett Z.D., Day N.F., Kimball T.K., Aamodt C.M., Heston J.B., Hilliard A.T., Xiao X., White S.A. FoxP2 isoforms delineate spatiotemporal transcriptional networks for vocal learning in the zebra finch. We thank Jennifer Morales and Maria Truong for their assistance in analyzing song behavior.

Chapter 3 is a version of Burkett Z.D., Day N.F., Peñagarikano O., Geschwind D.H., White S.A. VoICE: A semi-automated pipeline for standardizing vocal analysis across models. *Scientific Reports,* 2015; 5:10237. doi: 10.1038/srep10237. We thank AT Hilliard (Stanford) and S Horvath (University of California, Los Angeles (UCLA)) for consultation on the application of clustering methods, A Garfinkel (UCLA) for statistical advice, and ER Fraley (UCLA) for feedback about mouse vocalizations. We also thank BS Abrahams (Albert Einstein College of Medicine) for collecting pilot data.

Appendix 1 is a version of Chen Q., Burkett Z.D., Heston J.B., and White S.A. Expression Analysis of Speech-Related Genes FoxP1 and FoxP2 and Their Relation to Singing Behavior in Two Songbird Species. *The Journal of Experimental Biology,* 216, 3682-3689. Lily Sung assisted in the preparation of brain sections. The authors thank Dr. Julie E. Miller and two anonymous reviewers for their constructive comments on the manuscript.

Appendix 2 is a version of Miller J.E., Hafzalla G.H., Burkett Z.D., Fox C.M., White S.A. (2015) Reduced vocal variability in a zebra finch model of dopamine depletion: implications for Parkinson's disease. *Physiological Reports*, **3** (11), e12599, doi: 10.14814/phy2.12599. We thank Debora Lee and Jingwen Yao (UCLA) for behavioral analyses, Jonathan B. Heston (UCLA) for input on the figures; Stephanie Munger, Joshua L. Ritter, Patty Jansma, and Professor Nicholas Strausfeld (U. Arizona) for technical contributions.

Appendix 3 is a version of Fraley E.R., Burkett Z.D., Day N.F., Schwartz B.A., Phelps P.E., White S.A. (2016) Mice with Dab1 or Vldlr insufficiency exhibit abnormal neonatal vocalization

patterns. *Scientific Reports* **6**, 25807, doi: 10.1038/srep25807. We thank Ava DeLu for help in processing USVs, and Aly Mulji for assistance with mouse breeding and care.

VITA

2008	B.S., Zoology California State University, Long Beach
2011	M.S., Physiological Science University of California, Los Angeles
2011-2013	Molecular, Cellular, and Integrative Physiology Training Grant University of California, Los Angeles
2013	Eureka Scholarship Department of Integrative Biology & Physiology University of California, Los Angeles
2014	Edith Hyde Graduate Fellowship Department of Integrative Biology & Physiology University of California, Los Angeles
2015-2016	Neuroendocrinology, Sex Differences, and Reproduction Training Grant University of California, Los Angeles

Publications

Chen Q, Burkett ZD, Heston JB, White SA. 2013. Expression analysis of speech-related genes *FoxP1* and *FoxP2* and their relation to singing behavior in two songbird species. J Exp Biol 216:3682-3689.

Burkett ZD, Day NF, Peñagarikano O., Geschwind DH, White SA. 2015. VoICE: A semi-automated pipeline for standardizing vocal analysis across models. Sci Rep: 5:10237.

Miller JE, Hafzalla GH, Burkett ZD, Fox CM, White SA. 2015. Reduced vocal variability in a zebra finch model of dopamine depletion: implications for Parkinson's disease. Physiol Rep 3:e12599.

Berg JM, Lee C, Chen L, Galvan L, Cepeda C, Chen JY, Peñagarikano O., Stein JL, Li A, Oguro-Ando A, Miller JA, Vashisht AA, Starks ME, Kite EP, Tam E, Al-Sharif NB, Burkett ZD, White SA, Fears SC, Levine MS, Wohlschlegel JA, Geschwind DH. 2015. JAKMIP1, a Novel Regulator of Neuronal Translation, Modulates Synaptic Function and Autistic-like Behaviors in Mouse. Neuron 88:1137-91.

Fraley ER, Burkett ZD, Day NF, Schwartz BA, Phelps PE, White SA. 2016. Mice with Dab1 or Vldlr insufficiency exhibit abnormal neonatal vocalization patterns. Sci Rep 6:25807.

Burkett ZD, Day NF, Hilliard AT, Kimball TK, Aamodt CM, Heston JB, Xiao X, White SA. FoxP2 isoforms delineate spatiotemporal transcriptional networks for vocal learning in the zebra finch. (In preparation.)

Chapter 1: Introduction

The zebra finch songbird (*Taeniopygia guttata*) is a classical ethological model for studying learned vocal behavior because unlike most other laboratory models, they learn their vocalizations. The presence of discrete and interconnected groups of cells devoted to the learning and/or production of songs, referred to as the song circuit, present an attractive and accessible system in which to investigate the molecules that are linked to vocal behavior. This system has been used for decades to interrogate genes that are specifically related to vocalization. Significant reviews of the neurogenomics underlying songbird learning and memory already exist, though these reviews primarily focus on the auditory forebrain in which learning related transcriptional patterns occur following exposure to novel songs, referred to as song recognition learning [1]. The work summarized therein provides significant insight to the molecular cascades that are initiated in the process of experiencing a novel song, a prerequisite for developing a song template and initiating the learning process.

While necessary for producing and maintaining learned vocalizations, focus on the auditory forebrain excludes the telencephalic song learning and production circuitry, which is necessary for learning to produce a song, referred to as vocal production learning (VPL). VPL is viewed as a classical implementation of trial and error based motor learning wherein an animal in an active learning state practices its song while evaluating it against a stored template. In this introductory chapter, I first introduce the songbird as the premier model system for studying the molecular basis of learned vocalization. I then summarize the past few decades' work in describing genes regulated in the avian brain by vocal behavior and how those genes have shaped our understanding of the behavior, with emphasis on VPL. I then highlight contemporary techniques

that move beyond traditional differential expression based analysis to emphasize the context in which genes express as an important consideration when associating genes with behavior.

Songbirds as a model for the genetic basis of learned vocalization

Language is a highly-specialized behavior that seemingly is specific to humans, making its molecular determinants simultaneously intriguing and difficult to model in the laboratory. Language learning can be broken down into subcomponents, one of which – vocal learning – is shared by a moderate number of animal taxa, including humans, three avian lineages (parrots, hummingbirds, and oscine songbirds), and some members of the cetacean, pinniped, and pachyderm lineages. Among these, songbirds, particularly the zebra finch, have proven the most feasible for study in the laboratory and their anatomy, behavior, and genetics indicate a strong analogy with humans.

The zebra finch song circuit is a well-defined and interconnected series of clustered cells (referred to as song control nuclei) that form two distinct but connected pathways (Figure 1-1A). The posterior vocal pathway consists of a projection from cortical HVC (proper name) to the robust nucleus of the arcopallium (RA), which then connects to the motor neurons of the 12th tracheosyringeal nucleus that innervate the vocal organs. Lesions to the nuclei of this pathway eliminate the bird's ability to sing, indicating that it is primarily devoted to producing learned vocalizations [2]. A second pathway, the anterior forebrain pathway, is necessary for learning vocalizations. This pathway contains projections from striatopallidal Area X to the dorsolateral nucleus of the thalamus (DLM), to the lateral magnocellular nucleus of the anterior nidopallium

(LMAN). LMAN then projects both to Area X and provides input to the motor pathway by projecting to RA where it modulates vocal output [3]. Lesions to the AFP affect learning and variability in acoustic output [4-6]. The primary focus of this dissertation is Area X, which receives input from both HVC and LMAN, then projects to DLM. Area X is a basal ganglia homolog whose main output from medium spiny neurons is inhibition of DLM [7-9], a thalamocortical circuit homologous to that present in mammals. Recent work has shown that similar gene expression patterns are observed in humans and zebra finch brains that "specialize" regions for vocalization from their surrounding area (Figure 1-1B) [10].

In addition to the well-defined and analogous to human anatomy of the songbird brain, numerous behavioral parallels with humans exist. In both species, the vocal learning process begins with a period of sensory learning, wherein the learner is listening and forming neural template for the sounds that they will eventually learn. This period elapses from ~18d to 65d in zebra finches. This period of sensory acquisition is proceeded by and overlaps with a sensorimotor period, wherein the young finch begins singing "subsong", an avian form of babbling. Over the course of the next ~30 days, the juvenile finch will undergo trial and error learning in matching the song template acquired during the sensory period. At ~90d, the finch reaches sexual maturity and the vocalizations are "crystallized" such that subsequent broad changes do not occur. While this differs somewhat with speech as humans are capable of adding new words to their vocabularies throughout life, the capacity to learn specific phonemes and new languages is diminished at puberty (Figure 1-2).

Songbirds and humans also share similar reliance on a handful of genes of known importance in vocal learning, most notably FoxP2, whose mutation causes a rare and heritable speech and language disorder known as developmental verbal dyspraxia (DVD). DVD was first described in the KE family [11,12]. This disorder is primarily characterized by deficiency in planning and executing the sequence of movement of the organs required for producing speech [12]. The genetic basis for DVD was linked to a mutation in the DNA binding domain of FOXP2, where a point mutation causing an arginine to histidine substitution at position 553 gives rise to the KE family phenotype [11]. Despite their speech and language phenotype, affected members of the KE family have non-verbal intelligence quotient scores not far outside the population mean, indicating that cognitive deficits were speech-centric and not due to more general cognitive deficiency [12].

The link between FOXP2 and a speech and language disorder earned it the moniker of the "speech gene" and prompted decades' worth of research into the molecule. Being a transcription factor, FOXP2's function in gene regulation was an obvious line of inquiry. Numerous targets were identified in human [13] and mouse [14] and posited genes important for the molecular cascades underlying vocal learning. Studies in murine model systems have generated valuable insight to Foxp2's function, but the relationship between molecule and vocal learning behavior cannot be made in a vocal nonlearner.

Foundational studies in behavioral gene expression in the zebra finch brain

Early growth response gene-1 (*EGR1*) is a transcription factor that is broadly expressed across eukaryotic cells. It plays a role in a number of biological processes, including cell differentiation, inflammation, tumorigenesis, growth, and synaptic plasticity. *EGR1* expression is induced by depolarization of neurons, leading to its immediate transcription. Since *EGR1* plays a role in synaptic plasticity, it was an attractive target for study in the zebra finch. Early experiments revealed that *EGR1* expression is induced in the auditory forebrain of zebra finches and canaries while listening to heterospecfic songs [15], providing foundational evidence that behaviorally-relevant gene expression occurs in the songbird brain. Repeated exposure to the same song resulted in attenuated *EGR1* expression whereas exposure to novel songs induced it, providing a basis for cellular memory of a specific song [16]. Notably, these experiments observed no expression of *EGR1* throughout the circuitry for the learning or production of song, providing no indication as to its function (if any) in the execution of a behavior despite auditory responsiveness of the song circuit [17-24].

The link between the gene and song production was made when *EGR1* expression was observed throughout the song circuit in HVC, Area X, LMAN, and RA proportionally with singing, providing the initial evidence of gene expression driven by vocal behavior. Up to 60-fold induction over baseline levels were observed after 30 minutes of vigorous singing [25]. *EGR1* expression peaks following an hour's continuous singing, at which point it declines but maintains elevation beyond basal levels. Notably, the expression of *EGR1* was attributed solely to motor action and not song perception as birds that were deafened or muted were capable of inducing

EGR1 expression by performing song behavior in the absence of auditory input. This behaviorally driven expression of *EGR1* was observed in all of the song nuclei of both juveniles and adults except for in nucleus RA, in which *EGR1* upregulation occurred with singing in juveniles but was confined to a subsection of RA in adults, prompting hypotheses that diminished *EGR1* response in RA denotes the transition from plastic to crystallized song [26].

Following the discovery of the song-induced *EGR1*, other IEGs were found to have a similar behavior-expression relationship in all or portions of the auditory and/or song learning and production circuitry. C-FOS (protein) was induced in HVC and RA by 50 minutes of female-induced singing behavior [27] but not elsewhere in the song system. Similar to the observations of *EGR1*, c-FOS expression occurred regardless of whether the animal was capable of hearing itself, providing strong evidence that expression of cFOS is motor driven. More recently, another IEG and transcription factor, *Arc*, was identified as regulated by behavior in the songbird forebrain and it was shown to colocalize with *EGR1* and *c-fos* [28]. These observations provided foundational evidence that plasticity related transcriptional patterns occur in cells activated by motor behavior, prompting new research into the genomics underlying the behavior.

BDNF and UCHL1

Brain derived neurotrophic factor (BDNF) is another molecule with strong associations to learning, memory, and development. BDNF is a member of the trophin family of growth factors, which have broad roles in the development of the nervous system and in the stimulation and control of neurogenesis in juvenile and adult brains. An extensive role in synaptic plasticity and learning has been ascribed to BDNF. The song nucleus HVC receives new neurons throughout life and, in

seasonal breeding birds like canaries, HVC volume and blood testosterone levels are positively correlated. These seasonal changes in testosterone levels were linked to a mechanism wherein BDNF mediates cell recruitment and survival in HVC through the TrkB receptor. BDNF expression in canary HVC is positively associated with singing and the incorporation and survival of new cells [29]. These observations comprise a mechanism wherein seasonal elevation in testosterone drives the impetus for the animal to sing which then promotes increased HVC volume through singing-dependent BDNF expression enhancing cellular migration and integration into the region.

A gene that codes for a protein involved in deubiqutination, ubiquitin carboxy-terminal hydroloase (UCHL1), is positively regulated by song in HVC [30] and its expression is significantly greater in cells that are not turned over in adulthood, serving as a molecular marker for cells that are not replaceable in the circuit.

FoxP2

With the discovery that a mutation in the DNA binding domain of FOXP2 caused a speech and language phenotype in humans, investigation of this gene in a vocal learning model system was the next logical step. Expression patterns of FoxP2 in the developing human and finch brain are similar [31]. As with the genes discussed above, *FoxP2* is negatively regulated by singing behavior in the striatopallidal song nucleus Area X [32]. This behavioral regulation of the mRNA is specific only to the bird singing in a state of vocal practice, termed undirected (UD) singing, and not during vocal performance (FD) to a female. It was later shown that FoxP2 protein is downregulated by singing regardless of the social context [33]. When this behavioral regulation is

broken throughout sensorimotor learning by knockdown [34] or overexpression [35], birds exhibit a similarly poor vocal learning phenotype. Overexpression also eliminates the ability of the bird to add variability to its song – likened to a diminished capacity for vocal-motor exploration – providing a mechanism in which a lack of variability prevents acquisition of a fine motor skill.

Beyond 'A Gene at a Time': High throughput approaches to songbird behavioral genomics

All of the genes discussed in above were identified as interesting targets for investigation using low throughput approaches or by having specific hypotheses regarding their relationship to vocalization and/or development of the nervous system. Complex phenotypes are not monogenetic in origin, prompting the necessity to move beyond the candidate gene approach described thus far and into an era where the expression patterns of many genes can be simultaneously measured, permitting a deductive methodology. Through the advent of microarrays and more recently through next generation sequencing technologies, experiments such as those described above can now be massively scaled up so as to provide high-throughput discovery of genes related to vocal behavior. In this section, I describe some of the foundational and more recent experiments by which our collective knowledge of behaviorally regulated gene expression has been greatly expanded.

Zebra finch microarrays and the genome assembly

A significant step forward was the development of zebra finch cDNA microarrays that allowed for quantification of ~5000 unique genes in the zebra finch brain [36,37]. For the first

time, high throughput interrogation of the zebra finch transcriptome was made possible. The first study to use this technology found 33 genes, including many of the ones described above, regulated by singing behavior (0.5-3 hours or singing) in some or all of the song control nuclei between. The cDNA microarrays in this study became part of a collaboration among three research groups that lead to the development of the zebra finch Agilent oligoarray [37-39], which became the standard for studying gene expression in the songbird brain moving forward.

These arrays were used to identify transcriptional patterns throughout the songbird brain, wherein thousands of genes in some or all of the song control nuclei were correlated to singing [40]. These transcriptional patterns were clustered into temporal and anatomical groups, creating a model for how cascades of gene expression occur as birds sing. Studies using microarrays were not limited to finding genes driven by behavior. Most notably they were used to identify transcriptional "specializations" common to the human and avian brain, where it was discovered that the evolutionarily distant brains use similar gene expression strategies to define vocalization-specific regions from their outlying areas [10].

The sequencing and assembly of the zebra finch genome was completed in 2010 [41], providing a resource that brought the songbird model into the next generation sequencing era. This technology has significant advantages compared to microarrays, most notably the lack of requirement for sequence or species-specific probes, ideal for nontraditional model systems. The zebra finch genome was assembled onto 33 chromosomes and 3 linkage groups and annotated by pipelines from NCBI, UCSC, and Ensembl. As a part of the study in which the genome was assembled, hundreds of additional genes in Area X that were positively or negatively correlated

to singing [41], including ncRNA, which have been posited to drive the complexity of higher organisms [42]. It was later shown in zebra finches that FoxP2 is regulated by multiple miRNA, providing evidence for this assertion [43].

WGCNA and transcriptional networks underlying vocalization

A unifying theme of the gene expression studies described above conducted in songbirds is their focus on amount of gene expression in a brain region and its correlation to a given behavior (e.g. singing). With the development of weighted gene coexpression network analysis (WGCNA; [44]), the focus shifted away from traditional microarray analyses where differential expression was emphasized and instead to grouping genes based on their coexpression patterns.

This approach uses the simple logic that if the expression of two genes changes together across multiple samples they likely share some functional relationship. WGCNA models the transcriptome as a scale-free network, a topology frequently observed in biological, computational, transportation, and social networks where there are many nodes that are poorly connected and few strongly connected "hub" nodes that are integral to network structure. A common real-world example of a scale-free network is the United States domestic airport map where a small number of hub airports receive the majority of air traffic.

In a gene coexpression network, the genes are nodes and their connection strengths are determined by how similarly correlated each gene-gene pair is across samples. Using the coexpression data, a topological overlap metric is calculated by examining the relationships not only between genes as pairs but also the relationships between the pair and their common network

neighbors. Topological overlap is then used to hierarchically cluster the genes and an automated algorithm is employed for gleaning groups from the resulting dendrogram [45]. The technique is referred to as "unsupervised" because only after genes are grouped into modules are traits considered, allowing gene expression and not the phenotype to drive the clustering.

By analyzing genes in correlation space instead of relative expression and grouping based on their coexpression relationships, the transcriptional context is considered. This transcriptional context is quantified in terms of "connectivity", the summed adjacency for the gene across the entire network or within its module. WGCNA has repeatedly shown predictive power in identifying molecules and pathways relevant to heart disease [46], autism [47,48], human cognitive evolution [49], cancer [50], and numerous other clinical traits.

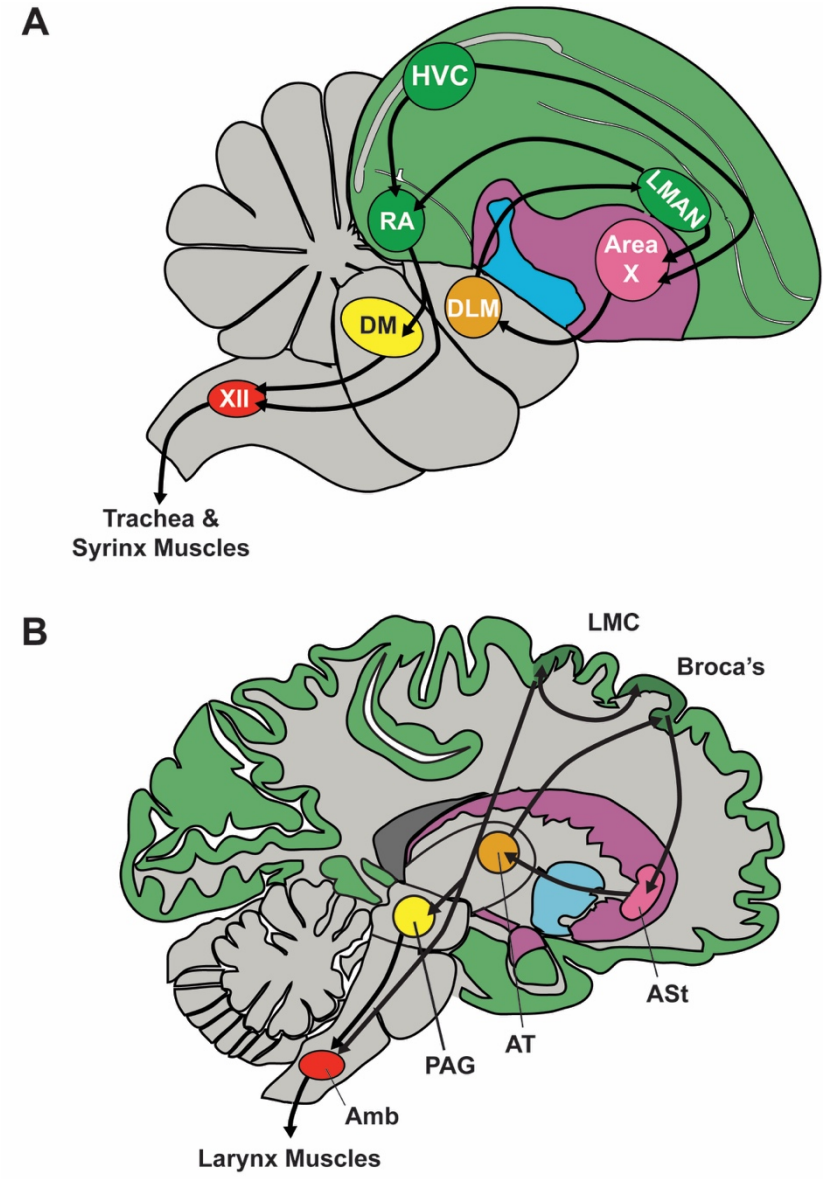
The first application of WGCNA to a naturally occurring behavior was done in our lab, where WGCNA was applied to the singing-induced transcriptome in Area X of adult songbirds [51]. This study served as a foundation for much of the work presented in this dissertation. The Agilent songbird microarray was used to quantify gene expression in Area X and the outlying non-song ventral striatopallidum (VSP) of adult birds across a continuum of singing states. The key findings were that ~1000 genes were either positively or negatively regulated by singing, including FoxP2. By emphasizing hub genes, novel pathways such as the Reelin signaling pathway were associated with singing behavior. The Reelin receptor VLDLR was a hub gene in a module negatively correlated to singing with significantly higher connectivity and stronger correlation to singing than the other module genes. In an illustration of the strength of the WGCNA approach, no difference in expression level between singing and non-singing birds was found. Thus, the

relationship between VLDLR and singing behavior is not apparent when examining how much the gene is expressed but only when considering the other genes with which it expresses. This principle was illustrated again when examining gene coexpression patterns in Area X vs VSP: absolute expression levels were tightly correlated between brain regions but connectivity was markedly different, indicating that the transcriptional context is the key feature in defining the singing-related brain region from the non-song brain region. These comparative network analyses allow for biological inferences to be made regarding the preservation or lack thereof in gene expression between two networks (Figure 1-3). In Chapter 2, I present evidence that learning-related gene coexpression patterns are present in both Area X and VSP and suggest that the overlap of singing- and learning-related gene coexpression that occurs only in Area X are the molecular building blocks of vocal learning in the basal ganglia.

Since the publication of our lab's work, a handful of groups have applied WGCNA to the avian transcriptome to better define the molecular determinants of zebra finch brain regions [52], territorial singing in white-crowned sparrows [53], and audition [54]. My work continues this trend in using network analysis to capture the coexpression patterns and genes most relevant to vocal learning.

Figures

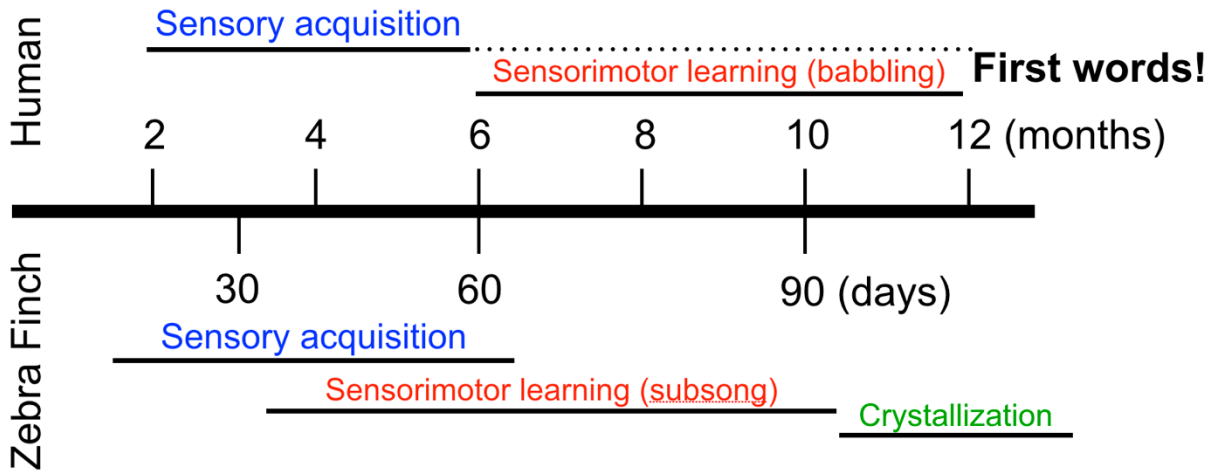
Figure 1-1: Songbird and human vocal neuroanatomy



a) The songbird brain consists of two pathways for learned vocalization. The song production pathway consists of projections from HVC (used as a proper name) to the robust nucleus of the arcopallium (RA). Projections from RA bifurcate to the dorsomedial nucleus of the midbrain

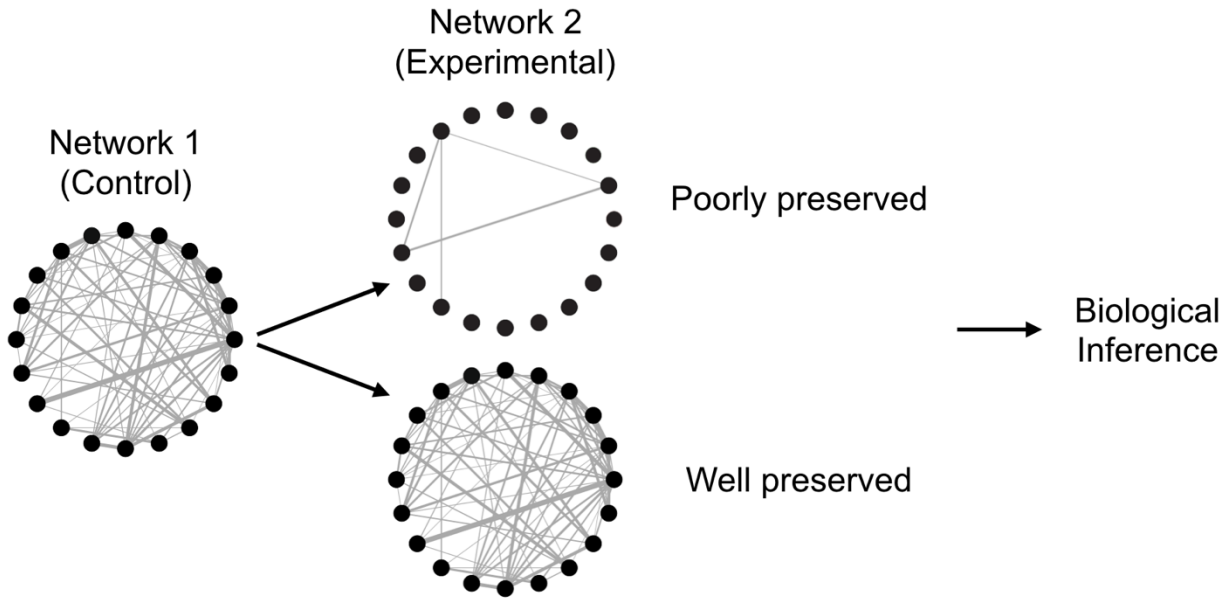
(DM) and the 12th tracheosyringeal motor neurons (XII), which innervate the vocal organs. The anterior forebrain pathway (AFP) consists of projections from Area X to the dorsolateral nucleus of the mesencephalon (DLM). DLM projects to the lateral magnocellular nucleus of the anterior nidopallium (LMAN), which sends projections to RA and back to Area X, forming a loop. HVC projects both to the song production and anterior forebrain pathway. **b)** The human brain contains analogous circuitry. Similar to the songbird AFP, projections from the anterior striatum (ASt) connect to the anterior thalamus (AT), which project to Broca's area. Broca's area then projects both to laryngeal motor cortex (LMC) and back to ASt, forming a loop. Like the song production pathway, the LMC projects to the nucleus ambiguous (Amb), which then innervates the vocal organs, a projection that is found only in vocal learning animals. In both panels, cortical regions are colored green, basal ganglia regions are colored pink, thalamic regions are colored orange, midbrain regions are colored yellow, and hindbrain regions are colored red. Figure is adapted from Arriaga et al., 2012.

Figure 1-2: Parallel developmental timelines between humans and zebra finches



The developmental timelines of human speech and zebra finch song both begin with a phase of purely perceptual sensory acquisition where the juvenile listens to conspecifics and forms a neural template of species-specific sounds (blue). In both species, this sensory phase is followed by a sensorimotor phase (red) where the juvenile begins making learned vocal utterances and begin refining them through trial and error learning. Upon reaching adulthood, zebra finch song is "crystallized" and unchanging whereas humans are capable of adding to their vocal repertoire throughout life. In both species, however, active maintenance of vocalization through hearing is required. Humans typically speak their first words around 12 months of age. Figure is adapted from Doupe and Kuhl, 1999.

Figure 1-3: Exemplar comparative network analysis



A key strength of WGCNA is the comparative network analysis, where the coexpression patterns between genes in a reference network ("Network 1") are compared to the coexpression patterns between genes in a test network ("Network 2"). How well these relationships are preserved across networks can then be used to make a biological inference about the two networks. This approach is utilized in Chapter 2 to understand how learning and singing related gene coexpression patterns are or are not altered by aging and/or anatomical brain region.

Chapter 2: FoxP2 Isoforms Delineate Spatiotemporal Transcriptional Networks for Vocal Learning in the Zebra Finch

Zachary D. Burkett, Nancy F. Day, Todd H. Kimball, Caitlin M. Aamodt, Jonathan B. Heston, Austin T. Hilliard, Xinshu Xiao, Stephanie A. White

Abstract

Human speech is one of the few examples of vocal learning among mammals, yet ~half of avian species exhibit this ability. Its genetic basis is unknown beyond a shared requirement for FoxP2 in both humans and zebra finches. Here we manipulated FoxP2 isoforms in Area X during a critical period for song development, delineating, for the first time, unique contributions of each to vocal learning. We used weighted gene coexpression network analysis of RNA-seq data to construct transcriptional profiles and found gene modules correlated to singing, learning, or vocal variability. The juvenile song modules were preserved adults, whereas the learning modules were not. The learning modules were preserved in the striatopallidum adjacent to Area X whereas the song modules were not. The confluence of learning and singing coexpression in juvenile, but not adult, Area X may underscore molecular processes that drive vocal learning in zebra finches and, by analogy, humans.

Introduction

The ability to learn new vocalizations is a key subcomponent of language. Complex behaviors such as human speech and birdsong typically involve suites of numerous interacting genes, making the attribution of their direct molecular underpinnings a challenge. While language is unique to humans, learned vocal behavior is present in a number of animal taxa. Among laboratory animals, a champion vocal learner is the zebra finch (*Taeniopygia guttata*), whose song learning shares numerous parallels with human speech development [55]. For example, both species share corticostriatal loops for producing vocalizations and have direct projections from cortical neurons onto brainstem motor neurons that control the vocal organs, a connection that is lacking or reduced in non-vocal learners [56-58]. In the brains of avian vocal learners, neurons controlling vocal production learning are uniquely clustered together within the surrounding corticostriatal circuitry, offering tractable targets for experimental manipulation. Despite their evolutionary distance, humans and zebra finches exhibit shared transcriptional profiles in key brain regions for vocal learning that are unique from surrounding brain areas and from the brains of non-vocal learning species [10].

The forkhead box P2 (FOXP2) transcription factor was the first molecule shown to be important for vocal learning in both humans and songbirds. Forkhead box proteins are characterized by the presence of DNA-binding FOX domains [59] and FOXP subfamily members form homo or heterodimers at zinc finger and leucine zipper domains in order to bind DNA. In humans, a heterozygous mutation in the FOX domain of FOXP2 causes a rare heritable speech and language disorder in a cohort known as the KE family [11,60], potentially by altering the

subcellular localization of the molecule [61]. While the mutation clearly disrupts vocal learning, multiple FOXP2 isoforms are endogenous to both songbirds and humans, including one that lacks the DNA binding domain. This truncated variant is referred to as FOXP2.10+ because, although it lacks the FOX domain, it retains the dimerization domains plus an additional 10 amino acids.

Consistent with its lack of a FOX domain, in vitro assays of FOXP2.10+ indicated that it does not localize to the nucleus [62]. Since it retains the dimerization domain, it has been hypothesized to act as a cytoplasmic sink, binding to other FOXP proteins and preventing their entry to the nucleus and interaction with DNA. Investigation of FoxP2 function in zebra finches has revealed remarkable parallels with humans. Both the full-length (FoxP2.FL) and FoxP2.10+ isoforms are present in each species [32], and similar *FoxP2* expression patterns occur in developing human and zebra finch brains [31]. In zebra finches, knockdown of *FoxP2* in the song dedicated striatopallidal nucleus, Area X, during vocal development impaired vocal mimicry of tutor songs [34], much as the KE family mutation impairs speech. These observations indicate that functional FoxP2 is necessary for proper vocal learning in both species.

The unique organization of song control circuit neurons enabled the discovery that FoxP2 is dynamically downregulated within Area X when zebra finches practice their songs, termed 'undirected' (UD) singing [32,33,63-65]. This FoxP2 decrease is accompanied by increased vocal variability [51,66], and blockade of FoxP2 downregulation impairs birds' ability to induce variability in their songs, thought to be a form of vocal exploration. A poor learning phenotype emerged following FoxP2 overexpression that was remarkably similar to that observed following knockdown [35]. Taken together, these results indicate that the dynamic regulation of at least

FoxP2.FL and thereby the behavior-linked up- and down-regulation of its transcriptional targets is necessary for the proper learning of vocalizations. No specific role in vocal behavior has yet been attributed to the FoxP2.10+ isoform.

These observations pinpoint FoxP2 as a molecular entry point to the pathways underlying vocal learning. In adult birds, we previously used Weighted Gene Coexpression Network Analysis (WGCNA) to identify thousands of genes regulated by singing specifically in Area X [51,67]). Since adult zebra finches sing stable, or crystallized, songs, the transcription patterns underlying vocal learning were not identified. Here we conduct a new study with two goals: 1) Determine whether FoxP2.10+ plays a role in vocalization and, 2) Manipulate FoxP2 isoforms in juveniles to generate a broad range of behavioral and transcriptional states upon which to apply WGCNA and thereby reveal learning-related gene modules. Toward the first goal, overexpression of FoxP2.10+ revealed a unique role for this truncated isoform in the acute modulation of vocal variability. Toward the second goal, overexpression of either GFP or one of the two FoxP2 isoforms created three distinct groups of juvenile birds: one that was good at learning and acutely modulating variability (GFP), one that was poor at learning and acutely modulating variability (FoxP2.FL), and one that was good at learning but injected stability into song (FoxP2.10+). We applied WGCNA to the Area X transcriptome of birds across this behavioral continuum and discovered striatopallidal coexpression patterns that were positively correlated to learning. These learning-related patterns were present in juvenile but not adult Area X. However, singing-driven coexpression patterns in Area X were largely preserved between juveniles and adults, suggesting that: 1) singing-related modules are independent of learning state and 2) the spatiotemporal co-

occurrence of both singing and learning-related gene modules in juvenile Area X is fundamental to vocal learning.

Methods

Subjects

All animal use was in accordance with NIH guidelines for experiments involving vertebrate animals and approved by the University of California, Los Angeles Chancellor's Institutional Animal Care and Use Committee. Birds were selected from breeding pairs in our colony.

Experimental Timeline

Our procedures closely followed those of Heston and White [35], which are detailed further in the Supplemental Experimental Procedures and schematized in Figure 2-2A. In brief, 30d male birds were injected with virus to overexpress GFP, FoxP2.FL, or FoxP2.10+ bilaterally in Area X, then recorded constantly until reaching 65d, when they were sacrificed by decapitation and brains rapidly frozen by liquid nitrogen. A total of 19 birds were injected (7 GFP, 6 FoxP2.FL, 6 FoxP2.10+). Sample sizes were selected so as to replicate the n used by Heston and White [35] where an n of 5 to 8 animals were required to observe a virus effect on tutor percentage similarity. In addition, the authors of the WGCNA R package recommend a minimum of 15 samples for building a network (https://labs.genetics.ucla.edu/horvath/CoexpressionNetwork/Rpackages/WGCNA/faq.html), so we ensured at least 5 animals per group. At ~60d, an NS-UD experiment was performed where an experimenter kept the animal from singing for the first two hours following lights-on to compare to the previous or following day's songs recorded after 2 hours of singing.

Surgery and Viruses

AAV1 identical to those used by Heston and White [35] and produced by Virovek (Hayward, CA) were used in this study. AAV1s expressing GFP, FoxP2.FL, or FoxP2.10+ behind the CMV early enhancer/chicken β actin (CAG) promoter were injected to Area X of 30d birds. Virus titers were all ~2.24E+13 vg/ml, thus equal volumes were delivered to each bird irrespective of the virus being injected. Immunostains (Figure 2-1D) were performed on tissue injected with a custom engineered HSV-1 expressing FoxP2.10+ behind the IE 4/5 promoter and GFP behind the CMV promoter. HSV was prepared by the virus core at The McGovern Institute for Brain Research at the Massachusetts Institute of Technology (Cambridge, MA).

Further information about the surgical procedure and viruses are in the Supplemental Experimental Procedures.

Song Analysis and Statistics

Motif Similarity: To quantify the acoustic similarity between pupil and tutor, a metric for quantity of learning, we utilized the Sound Analysis Pro (SAP) [68] Similarity Batch module. Asymmetric comparisons were performed between 10 tutor motifs from the final day before the pupil was isolated from the home cage and 20 pupil motifs every three days following the onset of singing in isolation following injection of virus. We used the average percentage similarity from these comparisons as a representative of how well the pupil had learned its tutor's song as of a given

analysis day. Statistical significance of motif similarity data was calculated by performing one-way ANOVA on the average percentage similarity score of each animal across virus groups within each time bin, as depicted in Figure 2-2D. If the ANOVA yielded a significant result, Tukey's Honest Significant Difference (HSD) was used as a post-hoc test.

Overall Vocal Variability: To broadly assess the amount of variability in the animal's song preceding sacrifice, asymmetric comparisons between 20 pupil motifs and themselves were conducted. We calculated the motif identity for all motif-motif comparisons as the product of their percentage similarity and accuracy divided by 100. Higher identity scores indicate lower variability within the batch.

Acute Vocal Variability Modulation: For finer-grained analyses of acoustic variability as presented in Figures 2-2C and 2-S1, we utilized SAP and VoICE [69] (https://github.com/zburkett/VoICE). Syllables from the first 20 minutes following two hours of non-singing or undirected singing on the NS-UD experiment days were hand segmented, had their acoustic features quantified in the SAP Feature Batch, then clustered by VoICE. Data for analyses of acoustic features were taken from the VoICE output. Effect sizes were calculated using the formula (NS-UD)/(NS+UD), where values were the CV of a given acoustic feature following two hours of NS or UD. Thus, negative values were indicative of increased song variability after UD singing (For more detail regarding this transformation, see Supplemental Information: Song Analysis). Statistical significance for each song feature was assessed by one-way ANOVA on the CV effect size for all syllables from all animals within each group. Tukey's HSD was used as a post-hoc test in the instance of a significant ANOVA result. For the raw acoustic data, as presented in Figure 2-S1, the syllables

were considered paired within virus construct and across singing context. Paired T-tests were used to assess whether two hours' silence vs. two hours' singing caused a significant change in the coefficient of variation for each acoustic feature.

Immunostaining

Tissue was prepared for immunostaining by sacrificing the animal 3-5 days following HSV injection then perfusing warm saline followed by ice cold 4% paraformaldehyde in 0.1 M phosphate buffer. Tissue was sectioned on a cryostat at 20 μM, thaw mounted on glass microscope slides, and stored at -80ºC. Thawed tissue was incubated with goat-anti-FoxP2 at 1:500 dilution (Abcam, Cambridge, UK [70]) and mouse-anti-Xpress at 1:500 dilution (ThermoFisher Scientific, Waltham, MA) overnight. AlexaFluor 546 donkey-anti-goat at 1:500 dilution and AlexaFluor 405 donkey-anti-mouse at 1:250 dilution secondary antibodies were used to bind the anti-FoxP2 and anti-Xpress signals, respectively. The tissue was visualized using a Zeiss (Oberkochen, Germany) LSM 800 confocal microscope and processed using NIH ImageJ[71].

Tissue Collection and Processing, RNA Extraction, cDNA Library Preparation, and Sequencing

Two hours following lights-on at ~65d, birds were sacrificed by decapitation. Brains were rapidly extracted and frozen on liquid nitrogen, then stored at -80ºC until all brains were collected. As in Hilliard et al. [51], tissue micropunches of Area X and VSP were performed. Brains were sectioned on a Cryostat until Area X became visible. Area X and outlying VSP were punched using a 20 gauge Luer adapter and stored in RNAlater (Qiagen, Germantown, MD) at -80ºC until RNA extraction was performed. Sections were then collected for validation of punch accuracy.

Total RNA extraction was performed in the same manner as in Hilliard et al. [51]. Samples in the present study were processed randomly and in parallel with another sequencing project. Tissue punches from both studies were processed in batches of 8. We used Qiagen RNeasy Micro Kits following the manufacturer's protocol and used QIAzol as the lysis reagent. We also performed an additional wash each in RW1 and RPE buffers beyond the manufacturer's protocol. Final elution volume was 20 μL. Extracted total RNA were stored at -80ºC until all RNA extractions were completed. All extractions were completed over the course of two weeks.

Total RNA was provided to the UCLA Neuroscience Genomics Core (UNGC; https://www.semel.ucla.edu/ungc) where RNA quality was assessed on an Agilent TapeStation (Agilent Technologies, Santa Clara, California). RNA of sufficient quality was then used to generate cDNA libraries using the Illumina TruSeq Stranded Poly-A Prep Kit (Illumina, San Diego, California). Libraries for each sample were divided across two lanes and sequenced in a total of 8 lanes in an Illumina HiSeq 2500 in high output mode, generating between 15 and 35 million 50bp paired-end reads per library.

RNA-Seq Preprocessing & WGCNA

FASTQ files for all 19 samples from UNGC were quality controlled and then aligned to the zebra finch genome assembly 3.2.4 (http://www.ncbi.nlm.nih.gov/assembly/524908) using Spliced Transcripts Alignment to a Reference (STAR) [72]. Reads mapping uniquely to exons were counted using the featureCounts() function in the Rsubread R package [73,74], then exon

counts were summed to the gene level and transformed to transcripts per million (TPM). Expression data were then log2 transformed and preprocessed for WGCNA as described in the Supplemental Experimental Procedures. Finally, we checked for batch effect on average expression resultant of RNA extraction group, RNA extraction experimenter, and across sequencing lanes. No batch effects were observed.

Signed weighted gene coexpression networks were constructed using the WGCNA R package [67] using custom written code for iteratively building networks, as described in the Supplemental Experimental Procedures. Soft thresholding power was set at 18 and 14 for the Area X and VSP networks, respectively. Minimum module sizes for both networks were set to 100 and the deepSplit parameters were set to 4 and 2 for Area X and VSP networks, respectively. All other input parameters were left at their default values.

Correlation of Behavior to Gene Expression

Calculation of gene significance to a trait requires the definition of a single value to which the amount of gene expression in each sample is correlated. Gene significances were calculated for the following traits: motifs, defined as the number of motifs each animal sang in the two hours following lights-on on the day of sacrifice; tutor similarity, defined as the percentage similarity between the pupil and its tutor on the day of sacrifice; variability induction, defined by inserting Wiener entropy CV scores into the equation (NS-UD)/(NS+UD) from the first twenty syllable renditions sung during the NS-UD experiment performed at ~60d; motif identity, defined as the product of the similarity and accuracy scores divided by 100 of the last 20 motifs sung by each

bird before sacrifice. Song variability was assessed on the motif level for the purpose of gene significance calculations so as to obtain a single value for each animal.

Following network construction, modules were summarized by calculating a module eigengene, defined as the first principal component of the module's expression data using the moduleEigengenes function in the WGCNA R package. The relationship between a module and a behavior was assessed by determining the Pearson correlation between the module eigengene and continuous behavioral traits as defined in 'Song Analysis and Statistics', above. Significance was then determined by calculating the Fisher transformation of each correlation using the corPvalueFisher() function in the WGCNA R package.

Gene Ontology, Module Significance, and Term Significance

At the time of our study, annotation of the zebra finch genome is relatively sparse, thus we used the HGNC gene symbols for the human homologs of the zebra finch genes in our study for gene ontology analyses. Genes with no known human homolog were excluded. Symbols were submitted to the GeneCards GeneAnalytics suite at http://geneanalytics.genecards.org [75]. GeneCards enrichment scores were converted into p-values, which were used as the input to module significance calculations. Module significance of a term was defined as the product of the average module membership for each gene annotated with a term and one minus the p-value for that term such that the genes with the highest module membership and lowest p-value prioritize the terms. Term significance was defined by weighting the module significance score by the gene significance for a given behavioral metric.

Transcription Factor Binding Site Analysis

The FoxP2 consensus binding sequence from the JASPAR database [76,77] was converted into a position-weight matrix (PWM) and used to scan the promoter (defined as the first 1000 base pairs upstream of the transcription start site) for each gene in the zebra finch genome. Putative FoxP2 binding sites were identified using the matchPWM function in the Biostrings R package (https://bioconductor.org/packages/release/bioc/html/Biostrings.html) with a minimum hit score of 80%.

Protein Interaction Networks and Scaling of Interaction Confidence Scores

STRING is a comprehensive database of known and predicted protein-protein interactions derived from experimental data, coexpression data, automated text mining, and also pulls information from other interaction databases. STRING accepts gene symbols as input, then mines for interactions between those genes and assigns a confidence score between 0 and 1 based on the evidence in the database for the genes' interaction. We submitted gene symbols for the human homologs of module members to STRING then operated on the highest confidence interactions ($\geq$ 0.9) in downstream analyses.

Interaction scores were scaled by different metrics to emphasize or deemphasize network position and/or relationship to behavior. Those metrics are:

1. The product of each gene's connectivity in juvenile Area X network: emphasizes interactions between the most connected genes in the juvenile network

2. The product of each gene's differential connectivity between juvenile and adult Area X networks: emphasizes interactions between genes that are of high network importance in juveniles but not adults

3. The product of each gene's gene significance for learning or singing: emphasizes interactions between genes that are strongly correlated to behavior independent of their connectivity

4. The product of each gene's connectivity and gene significance: emphasizes interactions between genes that are strongly correlated to behavior and of highly connected in the juvenile network

Network Visualization and Interactive Figures

Network plots presented in this manuscript were constructed using the freely available plotting software, Gephi (https://gephi.org), using edge lists prepared in R and exported in the .GEXF format. Interactive figures were exported from Gephi using the Sigma.js Exporter plugin (https://github.com/oxfordinternetinstitute/gephi-plugins). A more detailed description of their construction is presented in the Supplemental Experimental Procedures.

Accession Information

Raw and processed RNA-seq and behavioral data for each bird are available at the Gene Expression Omnibus (GEO; https://www.ncbi.nlm.nih.gov/geo/) at accession number GSE96843.

Results

Virus-mediated overexpression of FoxP2 isoforms affected song learning and/or vocal variability

Adeno-associated viral constructs were used to drive expression of FoxP2.FL or FoxP2.10+ in Area X of developing males. To verify isoform-specific overexpression, we used two riboprobes in *in situ* hybridization: one antisense to a region common to both transcripts (mid probe) and one antisense to a region near the 3' end of FoxP2.FL (3' probe; Figure 2-1A). Robust signals in the striatopallidum were observed in both hemispheres using the mid probe but only in the hemisphere injected with the FoxP2.FL construct using the 3' probe (Figure 2-1B). These results indicate that each viral construct over-expressed its *FoxP2* isoform and was thus suitable for bilateral injection into juvenile males at 35d. To quantify this, we performed qRT-PCR with a set of primers that amplifies a region common to both transcripts [34,78] and another set specific to the FoxP2.10+ (see Supplemental Experimental Procedures). The first primer set indicated that *FoxP2* levels were higher in birds injected with either construct relative to control levels. When quantified by the second primer set we found elevated PCR product only in the animals injected with the FoxP2.10+ construct (Figure 2-1C). Taken together, these results indicate that both constructs were effective in elevating levels of their encoded FoxP2 isoform throughout the 30d experimental period.

Overexpression of a tagged form of FoxP210+ in SH-SY5Y suggested that FoxP2.10+ acts as a posttranslational regulator of FoxP2.FL through heterodimerization and the formation of cytoplasmic aggresomes [61]. We thus examined the protein-level distribution of FoxP2.10+ and

FoxP2.FL in the finch striatopallidum following overexpression of an N-terminus Xpress tagged FoxP2.10+ linked to a GFP reporter (see Methods). Transfected cells shared the distinctive FoxP2.10+ staining pattern of aggresomes seen previously. In FoxP2+ cells that co-expressed the Xpress tag and GFP reporter, endogenous FoxP2.FL signal was interspersed among Xpress-positive puncta. [61] (Figure 2-1D).

We previously found that in unmanipulated birds, two hours of UD singing is sufficient to decrease Area X *FoxP2* mRNA (as measured by both the mid and 3' probes) and protein [32,33]. This decrease is accompanied by increased vocal variability [51,66] which is abolished by overexpression of FoxP2.FL [35]. These observations indicate that downregulation of full length FoxP2 is necessary for acute vocal variability but we did not directly manipulate FoxP2.10+. Here, we repeated our behavioral protocols to test for the induction of vocal variability and included the FoxP2.10+ injected animals (Figure 2-2A, 2-2B). To assess whether UD singing drove an increase in vocal variability, we quantified the effect of two hours' UD singing on the coefficient of variation (CV) of acoustic features in ~60d birds overexpressing GFP, FoxP2.FL, or FoxP2.10+. As predicted, GFP-expressing animals exhibited a negative effect size for most acoustic features and FoxP2.FL overexpression diminished these practice-induced changes in vocal variability, replicating our previous findings [35] (Figure 2-2C).

Unexpectedly, in animals overexpressing FoxP2.10+, song variability after two hours of UD singing was significantly *less than* that after two hours of NS for syllable duration, amplitude modulation, and Wiener entropy (Figure 2-2C). Thus, rather than increasing vocal variability, the act of UD singing promoted *invariability* when behavior-driven down-regulation of FoxP2.10+

was blocked. We also examined variability in the raw acoustic features of NS and UD song and found that expression of either FoxP2 isoform did not dramatically alter variability, indicating that the virus specifically affected the modulation of variability and not its overall level (Figure 2-S1 and Supplemental Experimental Procedures). Surprisingly, despite its acute effect on practice-induced changes in song variability, overexpression of FoxP2.10+ did not impair overall vocal learning (Figure 2-2D, 2-2E). These results suggest that the ability to modulate between relatively low and high variability states is important for proper vocal learning.

In sum, our AAV constructs allowed us to generate groups of animals in distinct states of vocal variability and learning. GFP-injected birds learned well and displayed singing-induced variability in the acoustic features of song. FoxP2.FL birds learned poorly and had no difference in their songs' acoustic variability following practice. FoxP2.10+ birds learned well but seemed to exist in a state where practice drives *in*variability in vocal acoustics. As such, a broad degree of both learning and variability induction existed across groups (Figure 2-2F). Next, we used these behavioral metrics as correlates to gene coexpression patterns to interrogate the transcriptional profiles underlying these traits.

Gene modules in juvenile Area X that correlate to vocal behavior were enriched for communication and intellectual disability risk genes

We used RNA-seq to quantify gene transcription in Area X of 65d juveniles overexpressing GFP, FoxP2.FL or FoxP2.10+, then used WGCNA to identify gene coexpression modules and link them to song learning. In our final network (see Supplemental Experimental Procedures), 7461 genes formed 21 modules (Figures 2-3A, 2-3B). A strength of WGCNA is that coexpression

networks are built in an unsupervised fashion blind to phenotypic trait data, thus making any correlations between coexpression modules and traits highly compelling. After network construction, we found significant correlations between module eigengenes and the following behaviors: tutor percentage similarity (i.e. vocal learning; modules: black, blue, darkgreen, orange, royalblue), number of motifs (i.e. amount of singing; modules: darkred, green, greenyellow), singing-induced acoustic variability (i.e. variability induction; modules: black, brown, darkgreen, darkgrey, magenta, orange, pink, purple, turquoise), and motif identity (i.e. overall vocal variability; module: darkgrey) ($0.00008 < p < 0.05$; Figure 2-3B). We examined all modules whose p-value was ≤ 0.05 and calculated the relationship between module membership and gene significance (For definitions of WGCNA and network terms, please see Supplemental Information: WGCNA and Network Terminology). For most modules, strong significant correlations were observed for each trait, indicating that the genes most representative of the module's overall expression profile were those most strongly related to the behavior (Figure 2-3C).

Connectivity is the core network concept and genes with high connectivity have the strongest coexpression relationships across the entire network, indicating greater importance to overall network structure and biological significance. The purple, green, and pink modules contained the most densely interconnected genes (Figure 2-S2), and were all significantly correlated to percentage similarity to tutor (green) or singing-induced variability (purple, pink) (Figure 2-3B, 2-3D). Altogether, these findings indicate that information about the relationships between molecular activity and behavior was embedded in the actual structure of the network, and that a gene's relationship to a module or a module's relationship to the network was predictive of

strong behavioral relevance. Therefore, we examined the most well-connected/hub genes within the context of their module (genes with the greatest intramodular connectivity, Figure 2-3G) or the entire network (genes with the greatest whole-network connectivity). We discovered that many of these hub genes were known risk genes for human disease (Figure 2-3G). For example, of the 7462 genes in our network, Fragile X Mental Retardation 1 (FMR1) had the third highest connectivity and was the most well connected member of the green module (Figure 2-3G). Deficiency in FMR1 gives rise to Fragile X Syndrome, a genetic disease with a multitude of symptoms including intellectual deficiency and speech and language impairment.

To attribute biological meaning to the modules, zebra finch gene symbols were converted to their Human Genome Organisation (HUGO) Gene Nomenclature Committee (HGNC) paralogs, then submitted to GeneAnalytics, a comprehensive tool for the contextualization of gene set data that integrates across multiple databases [75]. We calculated a module significance score for the resulting disease, gene ontology, and pathway annotations returned from GeneAnalytics. This was done by scaling the module membership of each gene that was annotated with a term. Specifically, the p-value for that term was subtracted from 1 such that the terms related to the genes most representative of the module received the highest scores [51]. The top five terms for the black song module (negatively correlated to singing), the brown module (positively correlated to variability induction and henceforth referred to as a 'variability induction module'), and green learning module (positively correlated to learning) are shown in Figure 2-3F. Since most modules contained hundreds of genes, prioritizing the ontology terms by the connectivity their annotated genes allowed genes with the greatest network importance (Figure 2-3F) to emphasize the terms with the greatest biological importance (Figure 2-3E).

Juvenile Area X modules for learning, but not singing, were preserved in ventral striatopallidum

To validate the specificity of the Area X modules to vocal behavior, we compared the Area X network to a network constructed from the adjacent unspecialized ventral striatopallidum (VSP) from the same animals. Area X and VSP networks were constructed using the genes that were common to both standalone networks. We hypothesized that all of the modules related to singing in Area X would have no correlation to behavior in VSP since, despite its close proximity and similar cell type composition, it is not linked into the song circuit. Moreover, a significant body of evidence suggests that the song circuit evolved as a specialization of existing motor circuitry [10,79-81]. As predicted, no module in the VSP network displayed any correlation to any of the singing or learning behaviors (Figure 2-4A). We calculated module preservation statistics between the two brain regions and observed that the song modules were among the most poorly preserved across the two networks (Figures 2-4B, 2-4C), further underscoring that Area X is specialized for song. This lack of preservation was not the product of differential gene expression between the two regions (Figure 2-4D, top) but instead reflected altered connectivity among similar genes (Figure 2-4D, bottom). In striking contrast to the song modules, the green learning module was very strongly preserved in VSP (Figure 2-4B, Figure 2-3B), indicating a generalized learning-related coexpression state exists in the juvenile striatopallidum that is specialized for singing in Area X.

Juvenile Area X modules for singing, but not learning, were preserved in adult Area X

To provide further context for the modules observed in our network and how they relate to learned vocalization, we compared them with prior data from adult zebra finch Area X [51]. Our

present network captures a point in zebra finch development when birds were actively learning how to improve their songs whereas in adulthood, the learning process has ended and adult songs are "crystallized". Contrasts between juvenile and adult networks highlight gene coexpression patterns that change between the two learning states, and inform their molecular underpinnings.

Our previous study in adults found multiple modules in Area X that were significantly correlated to singing crystallized songs. We reasoned that if very similar co-expression patterns were present in juveniles, then they would likely be unrelated to learning. In this case, the capacity to learn a song might be attributable to other genes and/or the relationships between them. To compare across studies, we first built new, separate networks for both age groups composed only of genes common to the two original networks and then computed gene significance scores for all genes in both networks. We found a remarkable correlation between gene significances to singing in juveniles and adults (Figure 2-5A), showing that genes in Area X shared similar relationships to singing, whether it be positive, negative, or nonexistent, independent of the animal's age and learning state. The replicated discovery of specific sets of song-related genes across studies and ages speaks to the profound effect that singing behavior has on gene transcription profiles within the song-dedicated basal ganglia.

We next calculated module preservation across the two studies, which assesses how well the coexpression relationships between genes persist across ages [82]. We observed strong relationships between module preservation and correlation to singing, and genes related to singing clustered together independent of age (Figures 2-5B, 2-5C). These results indicate that not only

are the relationships between genes and singing consistent across ages but those genes' coexpression patterns are preserved as well.

Since singing-driven gene coexpression patterns were similar between juvenile and adult Area X the capacity to learn vocalizations is not a product of large-scale differences in coexpression of the song module genes. We therefore looked for any modules that differed between juvenile and adult Area X. We found that the green, greenyellow and darkred learning modules that were significantly correlated to tutor similarity in juveniles were poorly preserved in adult Area X (Figures 2-5B, 2-5C). Like the genes in the song modules, learning module genes were activated with singing (Figure 2-5D, top row). For both well- and poorly-preserved modules, the ranked gene expression within each module displayed a positive correlation (Figure 2-5D, middle row). However, only the song modules showed positive correlations between connectivity in juvenile and adult Area X (Figure 2-5D, bottom row). These results attribute the difference between juvenile and adult Area X not to differential expression or altered correlation to behavior, but to differential connectivity in adults of modules that are correlated to tutor similarity in juveniles. Therefore, the capacity to alter vocalizations does not reside in the absolute expression level of a given gene but instead the transcriptional context in which that gene expresses. For example, FMR1 was poorly connected in the adult network but was positioned as a hub gene in the juvenile network, indicating the gene's importance during a developmental period when vocalizations are being actively modified but not during their maintenance.

A bioinformatics approach indicates MAPK11 as an entry point to neuromolecular networks for vocal learning

Above we described two classes of coexpression modules: 1) learning modules that were preserved throughout the striatopallidum but only present in juveniles, 2) song modules that were preserved across age but specific to Area X. Therefore, song modules and learning modules exist simultaneously only in juveniles, and their co-occurrence within Area X may reflect the capacity to dramatically alter vocalizations during sensorimotor learning. Therefore, we hypothesized that interactions between these two modules may drive the vocal learning process.

To test this idea using bioinformatics, we examined any genes linked to FoxP2, whose overexpression drove the broad range of tutor song copying in our animals. The gene with the greatest gene significance to learning was MAPK11 (Figure 2-6A). Interestingly, in Foxp2 heterozygous knockout mice, MAPK11 levels increase, supporting the interaction we observed here [83]. To test whether MAPK11 could be a target of FoxP2 in the zebra finch, we scanned the MAPK11 promoter for sequences corresponding to the FoxP2 binding motif from the JASPAR database (see Methods) [76,77]. We found a match with a single base difference beginning 288 base pairs upstream of the zebra finch MAPK11 transcription start site (Figure 2-6B). Taken together, these data suggest that birds overexpressing FoxP2.FL may be limited in their learning capacity to learn due to FoxP2 repression of this gene. In line with this, both the FoxP2.10+ and GFP animals had higher MAPK11 gene significances for tutor similarity than did FoxP2.FL animals (Figure 2-6A).

A strength of WGCNA is the "guilt by association" approach whereby genes in close network proximity to a gene of interest become candidates for a role in the same biological processes. With this in mind, we used MAPK11 as an entry point to pathways related to vocal

learning. We first scanned for genes with high topological overlap with MAPK11 (e.g. the closest network neighbors to MAPK11). Many of these genes were well-connected members of the green learning module (Figure 2-6C). One such gene, ATF2, had the fifth highest green intramodular connectivity and third highest whole network connectivity. ATF2 protein is necessary for proper development of the nervous system [84] and serves a dual purpose in affecting transcription by binding to cAMP response elements and also by acetylating histones H2B and H4 [85,86]. Like FMR1, ATF2 is poorly connected in the adult network [51].

While its role in development of the nervous system has been defined, no specific relationship between ATF2 and learned vocalization has been described. In our network, the ATF2 acetylation target histone H2B sorted into the blue module, which is strongly and positively correlated to the act of singing (Figure 2-3B) and has been linked to learning and memory in rat hippocampus [87]. These observations illustrate a pathway wherein overexpression of FoxP2 represses the expression of its putative target, MAPK11. As a consequence, less MAPK11 phosphorylation of learning module hub gene ATF2 would occur, decreasing acetylation of song module member histone H2B. This pathway represents an interaction between a network hub in a learning module (ATF2) and a song module gene (histone H2B) at a developmental time point at which the bird is actively learning its vocalizations.

To generalize this strategy, we used the Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) database [88] examine additional interactions between learning-related network hubs and singing genes in Area X. We submitted genes from the green, greenyellow, and darkred learning modules and the black, blue, darkgreen, orange, and royalblue song modules,

then filtered for cross-module interactions and scaled the confidence scores by the average intramodular connectivity of each gene in the interaction. This yielded a ranked list of interactions between genes positively regulated by learning and those regulated by singing, which was prioritized by weighted confidence score to yield the highest confidence interactions between genes with the greatest network importance. These interactions were visualized as a network (Figure 2-7). This approach allowed us to not only visualize the confidence in gene interactions but also the local neighborhoods formed by the protein interaction network, providing emphasis on genes of potentially greater importance in the vocal learning process based on the number of interactions they have.

We ranked interactions by four different metrics designed to emphasize or deemphasize gene significance, intramodular connectivity, and differential connectivity in juveniles vs. adults (see Methods). These metrics provide a basis for selecting protein-protein interactions based on the relationship to the genes and their most strongly correlated behavior, the coexpression network importance of the genes, or the change in connectivity between juvenile and adult birds. In using the latter metric, the decreased connectivity of learning-related genes ATF2 and FMR1 in adulthood is accounted for and interactions involving those genes are prioritized. Interactions between ATF2 and IRF2, DUSP5, and FOS are among the highest scoring interactions using this metric.

Discussion

In this study, we overexpressed FoxP2 isoforms to create groups of birds across a continuum of learning and ability to induce variability in their songs (Figure 2-2F), ideal for transcriptome profiling and WGCNA. We constructed an Area X gene network and discovered modules related to singing, learning, and vocal variability. The network properties of these modules revealed strong relationships between gene module membership and the behavior(s) to which the modules correlated.

To understand how gene coexpression patterns change across the boundary of the critical period for vocal learning, we performed comparisons between the network constructed here in juvenile Area X to one constructed from adult Area X [51]. We had competing hypotheses about whether the inability to learn new songs as an adult is resultant of changes to the song modules observed in juveniles or whether it is from some other transcriptional change. Module preservation statistics revealed very strong preservation of the juvenile Area X song modules in the adult network, supporting the latter hypothesis. In striking contrast, the densely interconnected juvenile striatopallidal green learning module was poorly preserved in adults, indicating that at least part of the learning related transcriptome is altered by aging.

Because we also created networks from VSP of the same animals, we were able to compare how well the Area X modules were preserved in a similar brain region lacking the specialization for singing. As observed in Hilliard et al. [51], Area X song modules were poorly preserved in VSP. In contrast, the densely interconnected green module was strongly preserved in VSP,

suggesting that a learning related gene coexpression pattern generalizes across the juvenile striatopallidum. These experiments define juvenile Area X as a nexus wherein the striatopallidal learning-related modules exist in tandem with singing-specific modules. As the brain ages, singing continues to drive transcriptional patterns but the learning related patterns are lost (Figure 2-8A). This hypothesis is supported by the preservation of the juvenile Area X song modules in adult Area X and preservation of the juvenile Area X green learning module in the outlying VSP (Figure 2-8B). Our findings suggest a model for the molecular basis for complex learned vocal behavior as not specific genes or coexpression modules, but the spatiotemporal combination of "singing" and "learning" building blocks that we observed in Area X. Like Area X, the other song nuclei of the finch brain likely evolved as specializations of existing motor circuitry [10,79]. We expect a similar principle to exist across the songbird brain where nonspecialized/learning related and specialized/singing related coexpression patterns converge to permit song learning.

Our findings here validate prior results where overexpression of FoxP2.FL made birds unable to induce variability into their songs and poorly learned their tutors' songs. This result supports the hypothesis that the behavior linked cycling of FoxP2 instead of the absolute level *per se* is critical for driving the vocal learning process. In addition, we described a behavioral role for the FoxP2.10+ isoform as we observed singing induced vocal *in*variability following its overexpression. A similar phenomenon was observed in a different species of passerine songbird, the Bengalese finch (*Lonchura striata domestica*), where two hours' UD singing resulted in less variable songs than those after two hours' NS [89]. In both species, the inability to induce variability into song did not affect vocal learning, suggesting that the ability to have relatively low

or high variability states in singing are necessary to properly learn a song regardless of whether those differential variability states precede or follow singing.

WGCNA identified *Fmr1* as a gene of great importance in a learning module. FMR protein is expressed throughout the zebra finch song circuit primarily in neurons, and the song system itself has been suggested as an interesting model system within which to study the gene's function [90,91]. FMR1 codes for an RNA-binding protein and therefore its level of expression could have a profound effect on a number of targets in the network [92]. Here, we observed a correlative link between FMR1 expression and how well the animal copied its tutor's song, a novel association that could be reasonably hypothesized given the speech and language phenotype concurrent with deficiency of the gene in humans. A key strength of WGCNA is the ability to query the network around genes known to be associated with a trait. FMR1's close network neighbors included ATF2, which is associated with learning but has no prior link to vocal behavior. We believe that further investigation into learning modules is likely to reveal molecules that are fundamental to learning behavior.

To identify those molecules that may interact at this particular developmental time point and brain region, we selected MAPK11 – a likely FoxP2 target [83] and the gene with the greatest significance to learning – to further investigate as an entry point to the pathways underlying learning behavior. Local neighborhood analysis of MAPK11 in the coexpression network revealed high topological overlap with many strongly connected members of green module, including the hub gene ATF2. ATF2 is a phosphorylation target of MAPK11 and this phosphorylation enhances its histone-acetyltransferase activity [93,94]. A known enzymatic substrate of ATF2 is histone

H2B [86], a member of the blue module that is positively correlated to singing. To probe for additional protein-protein interactions such as these, we mined the STRING database using song and learning module members, then prioritized the interactions based on the network properties and/or behavioral significance of the input genes. A prioritized list of interactions and a complex network emerged, highlighting genes based on their coexpression network importance and/or the number of protein level interactions in the database (Figure 2-7).

While there are differences in overall gene expression between the juvenile and adult brain, the context within which genes express, i.e. their connectivity, is drastically altered, especially in the learning modules. Changes in connectivity are not necessarily indicative of changes in the absolute level of a gene's expression, as evidenced by the comparisons between Area X and VSP (Figure 2-4D) or juvenile and adult Area X (Figure 2-5D), where expression levels correlate positively but connectivity does not. These data support the idea that the *coexpression* patterns, and thereby the genes' connectivity and network importance, contribute to the transition from a state of learning to a state of non-learning.

In using connectivity as a measure of network importance and protein interaction as a measure of functional biological output, the protein interaction landscape underlying learned vocal behavior shifts across the two developmental time points analyzed here. The local interaction network around green module hub ATF2 (defined as all those neighbors within two steps and with high confidence of protein interaction) is composed of well-connected genes in the learning and song modules (Figure 2-8C, top). Moreover, the connections to learning related genes are inputs to well-connected network hubs. As the juvenile crosses over into adulthood, the connectivity of

many of the learning-related genes, including ATF2, dramatically decreases. As part of the same process, the adjacencies between genes in the interaction network shift such that a connection to a learning-related gene is no longer one with a hub (Figure 2-8C, bottom). This shift in network importance may present a pattern underlying song maintenance rather than song learning, and potentially the closure of the critical period in which the bird can change its song.

To understand the mechanisms underlying the transition between the two learning states, our data highlight the importance of the network position of a gene, beyond simply its absolute expression level at a given time. To enable vocal plasticity after the closing of a critical period, a goal critically relevant to social and communication disorders, manipulations that coordinate gene expression such that poorly connected genes are reestablished as network hubs are likely required. Using exogenous sources to attain these goals is methodologically complex, but the pathways prioritized and presented here provide a foundation for breaking down the components of vocal learning behavior.

In sum, we have described the transcriptome at a developmentally significant point in the vocal learning process and provided context for it in terms of aging and brain region specificity. We have also suggested numerous coexpression and protein level interactions that our data indicate are significant to vocal learning. Due to the large amount of data generated by this study, we have generated as a supplement to the figures in the manuscript, interactive graphics describing the coexpression and protein interaction networks and compiled descriptive statistics and have hosted them on our laboratory website (https://www.ibp.ucla.edu/research/white/genenetwork.html). We strongly encourage the reader to explore these datasets to mine for coexpression and protein

interactions among their genes of interest. Further investigation of these pathways in the zebra finch is necessary for confirming their validity and providing the molecule-to-behavior links suggested herein. By using a network based approach, we are able to prioritize the interactions between genes and identify pathways so as to begin the process of teasing out the complex interactions underlying a complicated behavior.

Supplemental Information

WGCNA and Network Terminology

WGCNA is a well-established technique for gleaning biologically relevant clusters of coexpressed and functionally related genes from microarray and sequencing data. WGCNA methods and terminology are summarized and defined in numerous manuscripts [44,51,95-98]. For the sake of completeness, we provide working definitions of network terms that we use in the main text of this manuscript. Definitions of greater detail are available in the manuscripts cited above.

- **Adjacency (a):** The first step of network construction is to generate an adjacency matrix where $A_{ij} = S_{ij}^{\beta}$, where i and j are genes, S is the expression correlation across samples, and β is an empirically derived power to which the correlation is raised such that the resulting network approximates a scale free topology.
- **Connectivity (k):** Connectivity is a measure of connectedness of a given gene, either in the context of its module (kIN) or the entire network (kTotal). Connectivity is defined as follows:

  $$k_i = \sum_{j=1}^{N} a_{ij}$$

  where i and j are genes, N is all of the genes in the module or network, and a is the adjacency between genes i and j.

- **Topological overlap:** Adjacency is transformed to topological overlap as a method of calculating the interconnectedness (or similarity) between two nodes. Topological overlap is defined as follows:

$$\omega_{ij} = \frac{l_{ij} + a_{ij}}{\min\{k_i, k_j\} + 1 - a_{ij}} \quad \text{and} \quad l_{ij} = \sum_{u \neq i,j} a_{iu} a_{uj}$$

, where u represents all genes besides i and j. A and k are defined above.

- **Gene significance:** The Pearson correlation between a gene's expression profile and, in our work, a given behavioral metric.
- **Module eigengene:** The first principal component of a module's gene expression profile, a method of summarizing an entire module in one vector.
- **Module membership:** The correlation between an individual gene expression profile and a module eigengene. Genes with high module membership tend to have high intramodular connectivity and are referred to as intramodular hubs. Of note, genes can have high module membership in more than one module.
- **Zsummary:** Along with median rank, a term for quantifying preservation of gene coexpression patterns between two independent datasets[82], such as between juvenile and adult Area X or juvenile Area X and juvenile VSP. Zsummary is a composite preservation score defined as the average of Zdensity and Zconnectivity, which assess the preservation of connection strength among network nodes (e.g. Are strongly connected nodes in one network also strongly connected in the other?) and the connectivity patterns between nodes (e.g. Do the patterns of connection between specific nodes exist in both networks?),

respectively, following permutation tests under the null hypothesis. Higher Zsummary scores indicate better preservation.

Supplemental Experimental Procedures

Song Recording

Birds were recorded constantly from the initial placement of their home cage into a sound attenuation chamber at ~20 d to sacrifice at 65d. Countryman EMW or Shure SM93 omnidirectional lavalier microphones were used. Sounds were digitized using PreSonus FirePod or PreSonus Audioboxes at a 44.1 kHz sampling rate and 24-bit depth. Recordings were managed by Sound Analysis Pro 2011 software (SAP, [68]).

Detailed Experiment Timeline

Following methods established by Heston and White [35], we isolated breeding cages that contained candidate experimental birds along with their parents and siblings when the juveniles reached ~20d. The breeding cages were recorded constantly upon isolation so as to capture tutor vocalizations and ensure juveniles did not sing prior to surgery. At 30d, juvenile males were bilaterally injected with AAV1 to overexpress either FoxP2.FL, FoxP2.10+, or GFP and returned to their breeding cages following surgery. At 40d, juvenile males were isolated from all other birds and recorded constantly. At ~60d, an 'NS-UD' experiment was performed following the methods of Miller et al., Chen et al., and Heston et al. [35,66,89] to assess the bird's ability to induce variability into its song resulting from practice. On the 'NS' day, for the first two hours following lights-on, birds were distracted if they attempted to sing. (Those that sang >10 motifs were excluded from that day's experiment). On the 'UD' day, birds were allowed to sing unrestricted for the first two hours following lights-on. The level of variability in the animal's songs

immediately following those two hours was then quantified. Birds were later sacrificed at 65d following two hours of unrestricted singing. In order to assure a broad range of song amounts immediately preceding sacrifice (and to thereby capture a range of singing-induced gene expression), we prevented one bird in the GFP group from singing during the two hours preceding sacrifice.

Stereotaxic Surgery and Viruses

As described in Heston et al.[35], 30d juvenile males were anesthetized using 2-4% isoflurane in pure oxygen and secured in a custom-built avian stereotaxic apparatus, then injected with virus to overexpress FoxP2.FL, FoxP2.10+, or GFP bilaterally into Area X at the following coordinates: 45º head angle, 5.15 mm rostral of the bifurcation of the midsaggital sinus, 1.60 mm lateral of the midline, and to a depth of 3.3 mm. Virus was injected via a Drummond Nanoject II through a glass microelectrode with inner diameter between 30 and 50 μM backfilled with mineral oil. Three 27.6 nL injections were performed with a 15 second wait between injections and a 10-minute wait before retraction of the electrode so as to minimize vacuum action pulling the virus away from the injection site. Incisions in the scalp were closed with Vetbond (3M, St. Paul, MN, USA) and the animals given oxygen for 1-2 minutes until alert, upon which they were returned to their home cages.

AAVs used in this study to overexpress FoxP2.FL and GFP were identical to those produced by Virovek and used by Heston et al[35]. The AAV used to overexpress FoxP2.10+ was otherwise identical to those viruses, except its gene product was the complete coding sequence for FoxP2.10+, first discovered and cloned in zebra finch by Teramitsu et al. [32] (Genbank Accession

Number DQ285023). All virus titers were 2.24E+13 vg/ml, thus equal volumes of delivery were used for each virus. Using this method, Heston et al. [35] estimated that 24±5.5% of neurons at the epicenter of the virus injection are transfected by the virus and that 96.7±1.7% of cells that are transfected are neurons. These transfection rates are sufficient to observe a behavioral effect of the virus and were thus used in the present study.

We also used HSV to express a form of FoxP2.10+ labeled tagged with an Xpress epitope and included a GFP transfection reporter. The limited cloning capacity of AAV precluded our ability to express a reporter gene in the viruses that we used for behavior and RNA-seq experiments. We opted not to include an epitope tag in the AAV to prevent conformational changes to the exogenous FoxP2.FL or FoxP2.10+ proteins that may affect their ability to dimerize and/or interact with DNA. FoxP2.10+ is identical to FoxP2.FL except for a 10 amino acid difference at its C-terminus. No antibody specific to this isoform exists, thus we used the Xpress tag. HSV has a considerably larger cloning capacity than AAV, which allowed us to include a GFP reporter that expresses as its own transcript independent of FoxP2.10+. Surgical procedures were identical to those performed with AAV except the virus was diluted to 60% in dPBS immediately preceding injection, per the recommendation of the manufacturer. HSV reaches peak expression much more rapidly than does AAV, thus birds injected with HSV were sacrificed 3-5 days post-injection [99].

In Situ Hybridization

In situ hybridizations were performed following the procedures of Jacobs et al. [100] and using the two probes antisense to different regions of the zebra finch FoxP2 mRNA transcript as described in Teramitsu et al. [31]. 20 μM thick sections were thaw mounted onto Superfrost Plus

microscope slides (ThermoFisher Scientific, Waltham, MA, USA), then postfixed with 4% paraformaldehyde in PBS, pH 7.4. Sections were hybridized with [^{33}P]UTP-labeled RNA probes.

PCR Primers

Due to the shared sequence of the FoxP2.FL and FoxP2.10+ transcripts, we were unable to measure FoxP2.FL independently of FoxP2.10+. The primer pair used for FoxP2.FL has been published previously to quantify FoxP2 knockdown [34,78], thus we also used it here. The forward sequence used was 5'-CCTGGCTGTGAAAGCGTTTG-3' and reverse 5'ATTTGCACCCGACACTGAGC-3'. We designed a primer pair for FoxP2.10+ using the NCBI Primer-BLAST tool [101]. The input sequence to Primer-BLAST was the FoxP2.10+ mRNA CDS (GenBank accession DQ285023.1). The forward primer sequence used was 5'-CGCGAACGTCTTCAAGCAAT-3' and the reverse sequence used was 5'-AAAGCAATATGCACTTACAGGTT-3'. Primer specificity was determined by obtaining a single peak in melting curve analysis and obtaining a single amplicon of predicted size following qPCR. GAPDH forward and reverse primers were 5'-AACCAGCCAAGTACGATGACAT-3' and 5'-CCATCAGCAGCAGCCTTCA-3', respectively.

qRT-PCR Experiments

200 ng of RNA from Area X punches was reverse transcribed into cDNA using the Bio-Rad iScript cDNA Synthesis Kit (Hercules, CA, USA). 25 µL qPCR reactions were assembled in MicroAmp Optical 96-Well Reaction Plates (ThermoFisher Scientific). Reaction components were 0.5 µL cDNA, 200 nM primers, 12.5 µL PowerUp SYBR Green Master Mix (ThermoFisher

Scientific), and 10.75 uL nuclease-free water. Cycling conditions were 50°C for 2 minutes, 95°C for 2 minutes, then 40 cycles of 95°C for 15 seconds and 60°C for 1 minute. A dissociation step of 95°C for 15 seconds, 60°C for 1 minute, 95°C for 15 seconds, and 60°C for 15 seconds was then performed. All reactions were run in triplicate and all samples for an individual animal were run together on the sample plate. FoxP2 expression was quantified relative to GAPDH and normalized to the GFP-injected animals using the $2^{-\Delta\,\Delta C}{}_{T}$ method [102].

Song Analysis: (NS-UD)/(NS+UD) Effect Size vs. Raw Acoustic Feature CV

The calculation of effect size was performed because it allows for comparison across virus groups instead of a series of paired comparisons within group [103]. The transformation normalizes acoustic features so that any observed changes are viewed in the context of the initial values. We present a hypothetical example in the table below where a change of 50 Hz for two syllables is assessed as of higher magnitude following the transformation that we applied for our song data:

Syllable A				**Syllable B**			
NS	UD	Raw Delta	(NS-UD)/(NS+UD)	NS	UD	Raw Delta	(NS-UD)/(NS+UD)
100 Hz	150 Hz	50 Hz	-0.2	500 Hz	550 Hz	50 Hz	-0.048

RNA-Seq Preprocessing & WGCNA

Raw FASTQ files furnished by UNGC were first quality controlled using FASTQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). FASTQC returned results indicating high quality across all bases in each read in each sample and no adapter contamination was detected, therefore we did not perform any filtration of the reads before alignment. Reads were aligned to the NCBI zebra finch genome assembly 3.2.4 (http://www.ncbi.nlm.nih.gov/assembly/524908/) using STAR [72]. Mismatch tolerance was two base pairs and only uniquely mapping reads were considered in downstream analyses. The featureCounts() function in the Rsubread R package was used to count all reads mapping within exon features, then all exon counts were summed to the gene level so that each gene had a single value of reads mapped to it [73,74]. Gene expression was then quantified by calculation of transcripts per million (TPM). TPM values were log2 transformed and genes with zero variance across samples were removed. We used an iterative process of removing gene expression data from single samples whose expression was greater than 2.5 SD of that gene's expression across all samples, repeating until no samples remained with expression greater than 2.5 SD away from the gene's average expression across all samples. Finally, we calculated the intrasample correlation (ISC) and used a hard cutoff of 2 SD away from the group ISC for removal of samples from the study. No sample in any group (Area X or VSP) was greater than 2 SD from the group ISC. Data were quantile normalized as the last step. Final data input to WGCNA was 13665 and 13781 genes for Area X and VSP networks, respectively, across 19 total samples.

We calculated the soft thresholding power for construction of the WGCNA adjacency matrix using the pickSoftThreshold function in the WGCNA R package at 18 for Area X and 14 for VSP. We then constructed a signed network using the blockwiseModules function in the WGCNA R package. For the Area X network, we used a minimum module size of 100 genes and deepSplit was set equal to 4 for Area X and 2 for VSP. Genes were required to have at least a connectivity of 0.3 with their module eigengene in order to remain a member of their module and the module 'core' (= minimum module size/3) needed to have a minimum eigengene connectivity of 0.5 for the module to not be disbanded. All other parameters were set to default. Networks were iteratively constructed with genes in the grey module removed from the expression data after each round of network building and module definition. The networks were considered final after no genes were placed into the grey module.

During network construction, FoxP2 was removed, presumably due to the lack of coexpression with other genes in the network resulting from virus-driven overexpression. Therefore, we added FoxP2's expression data back into the final network and it became the only gene in the grey module. Once coexpression modules were defined, we correlated vocal behavior to the module eigengenes. Since the grey module included only a single gene with no significant behavioral correlations, it was excluded from module-trait analyses.

Network Visualization and Interactive Figures

We have created interactive versions of many of the network plots in this manuscript (Figure 2-2F), all additional Area X modules (similar to Figure 2-2F but not presented in the manuscript), and the protein interaction network presented in Figure 2-7. They are hosted at our

laboratory website (https://www.ibp.ucla.edu/research/white/genenetwork.html) along with high resolution static PDF versions.

In weighted coexpression networks, each node (i.e. gene) is connected to every other node in the network, even if the weight of the edge (i.e. connection) is zero. Therefore, plots depicting nodes and their edges with other genes become exceedingly complicated and unintuitive if all nodes and edges are included. In an effort to sparsify the networks and present the most salient data, we removed edges and genes from the coexpression networks using the following workflow: first, remove ≤ 98% of edges, then remove all disconnected nodes, then remove all nodes that are not part of the network's main component (e.g. the largest group of connected nodes). The remaining nodes and edges were plotted.

In this manuscript, we present three types of network plots that look similar but convey different data. The three types are as follows:

1. The entire gene coexpression network, as in Figure 2-S2 and https://sites.google.com/a/g.ucla.edu/genenet/coexpressionnetwork. In these plots, the nodes represent genes and their colors represent the module assignment. Edges represent the adjacency between nodes and the edge color is a combination of the origin and target node colors. Due to the overwhelming number of edges in this network, the edge weights are scaled to minimize the range. Node size in this network is equivalent to the node's degree (e.g. the number of connections originating or terminating at that node) and the maximum node size is suppressed so as to provide maximal visual clarity.

2. Individual coexpression modules, as in Figure 2-2F and https://sites.google.com/a/g.ucla.edu/genenet/modules. These plots are similar to the preceding except that, potentially, more nodes are present in the module since the filtration procedures detailed above are applied in a different context (e.g. only the expression data in the module are considered here vs. the expression data for the entire network). The same scaling parameters as above are applied to the edges for visual clarity.

3. Protein interaction network, as in Figure 2-7 and https://sites.google.com/a/g.ucla.edu/genenet/protein. Nodes represent proteins and their colors represent the coexpression module assignments. Node size is equivalent to its degree. Here, the edge width conveys meaning and is helpful in interpreting the relationship between nodes. An edge is drawn between two nodes when the STRING database indicates a high confidence interaction (score ≥ 0.9) between them. Edge widths are the confidence score scaled by the product of the origin and target node's intramodular connectivities (kIN). Thus, thick edges indicate a high confidence protein level interaction between two genes that are well connected members of learning and singing related modules. Unlike the previous plots, a node's size does not necessarily convey a higher degree of coexpression network importance. Instead, it indicates many interactions involving this protein described in the database. The thickness of the edges conveys influence of the gene's biological importance, as interpreted through their kIN. Whether a node's degree or the weight of its connections is the ultimate determinant of its relationship to vocal learning remains to be

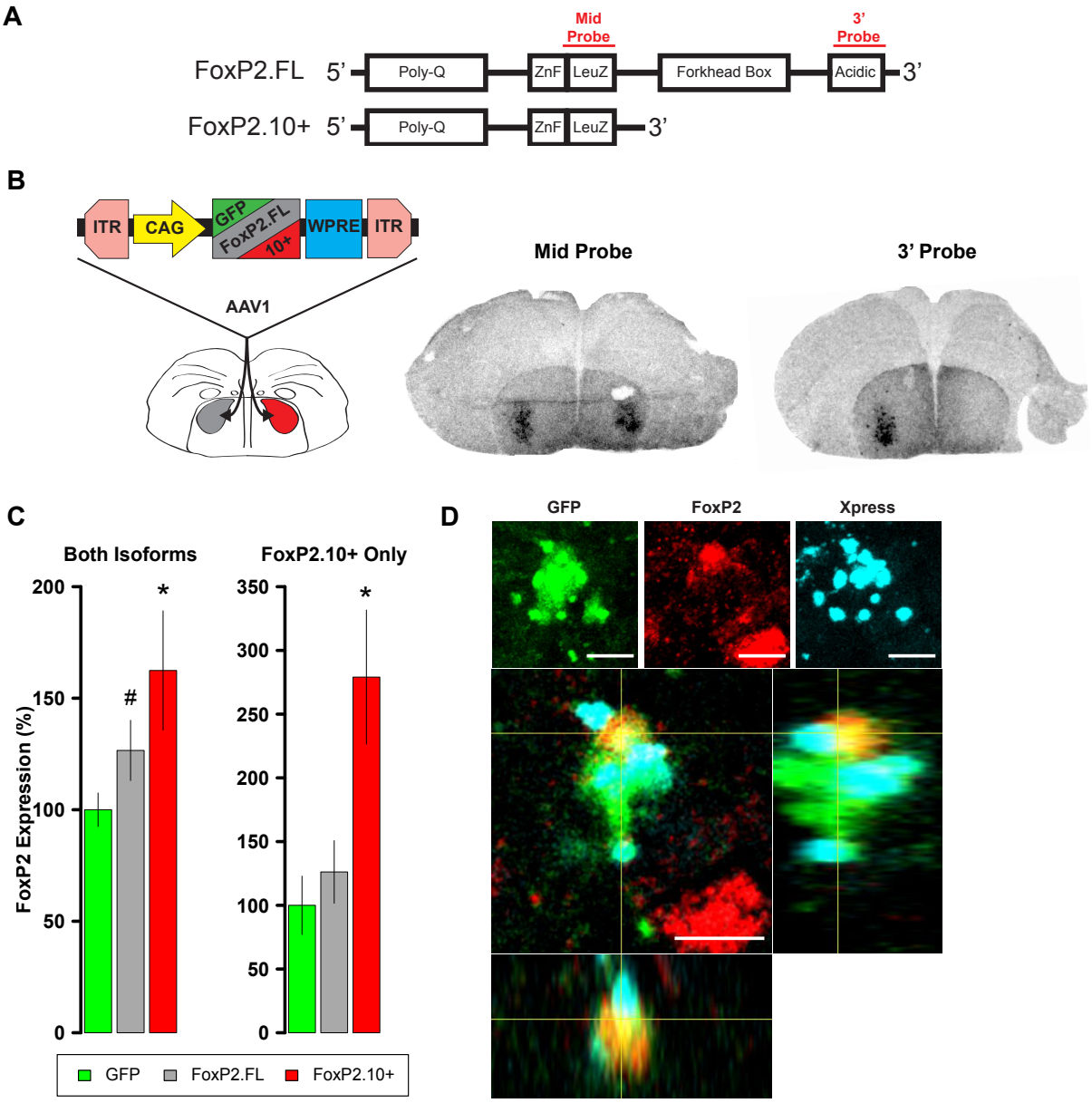
determined but the reader should keep the preceding information in mind when interpreting this network.

Acknowledgements

We thank Jennifer Morales and Maria Truong for their assistance in analyzing song behavior. This work was supported by NIH grant RO1MH07012 (SAW). ZDB received training funds from the "Neuroendocrinology, Sex Differences, and Reproduction" training grant 5T32HD007228.

Figures

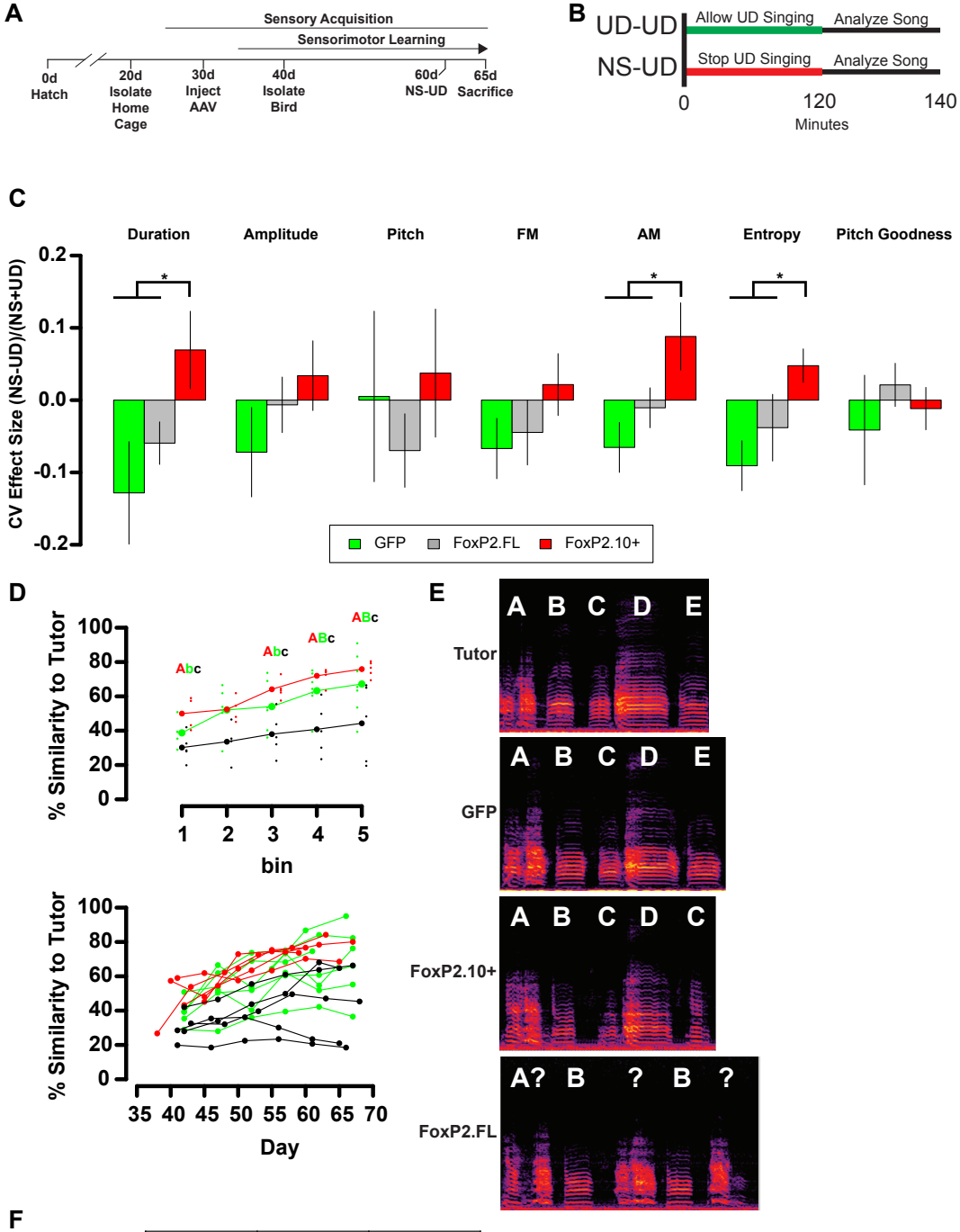
Figure 2-1: Overexpression of FoxP2 isoforms



a) Schematics of the full-length (FoxP2.FL) and 10+ (FoxP2.10+) transcripts. The regions targeted by the complementary RNA probes are shown in red. **b)** Left panel depicts experimental design to test for isoform-specific expression in vivo. Middle and right images depict two sections from the

same female brain. For purposes of validation only, the bird's right hemisphere (shown on left) was injected with an AAV expressing FoxP2.FL while the left hemisphere was injected with the FoxP2.10+ construct. Two weeks post-injection, robust signals in the striatopallidum were observed in both hemispheres using the mid probe but only in the hemisphere injected with the FoxP2.FL construct using the 3' probe. Signals reflect both the endogenous FoxP2 expression pattern [31,32,104] as well as enhanced levels due to viral-driven expression. **c)** FoxP2 expression quantified by qRT-PCR in juvenile males that were bilaterally injected with one of the constructs at 35d using primers that identify both isoforms (left graph) or the FoxP2.10+ isoform only (right graph). Enhanced expression is observed in the FoxP2.FL (grey; 126.5 ± 13.53%; n=6;) and FoxP2.10+ (red; 162.4 ± 26.77%; n=6) groups relative to levels of birds that received the GFP control construct (green; 100±7.54%; n=7). Values represent percentage relative to GFP ± SEM. * and # denote p = 0.0309 and p = 0.0841, respectively, of an unpaired two-tailed bootstrap vs. GFP. **d)** A cell in the zebra finch striatopallidum expressing GFP (indicating viral transfection; green), endogenous FoxP2 as revealed by an antibody directed to the C-terminus (red), and Xpress-FoxP2.10+ revealed by an antibody to the Xpress tag (cyan). The Xpress signal is reminiscent of FoxP2.10+ aggresomes observed by Vernes et al. [62].

Figure 2-2: Overexpression of FoxP2 isoforms affect learning and/or variability in song



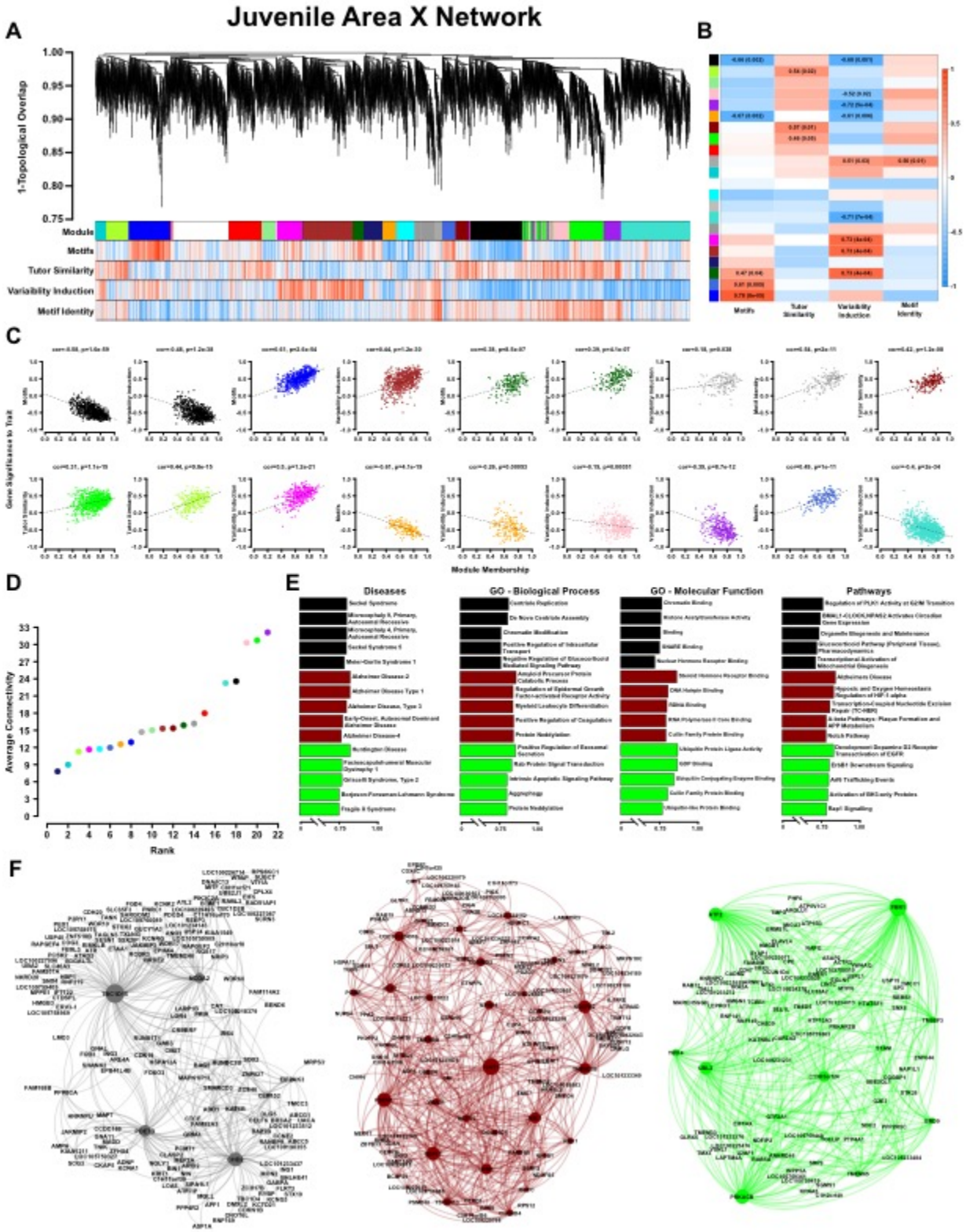
F

	GFP	FoxP2.FL	FoxP2.10+
Learning	Good	Poor	Good
Variability after UD	Increases	No change	Decreases

a) Timeline of experimental procedures relative to critical periods in song development. **b)** Schematic illustrates NS-UD or UD-UD experiments performed on adjacent days. **c)** The effect size of two hours' UD singing on syllable CV was calculated using the formula (NS-UD)/(NS+UD) after an NS-UD, UD-UD experiment performed at ~60d and 61d as in (b). Overexpression of FoxP2.FL (grey bars; n = 16 syllables; Duration = -0.059 ± 0.029; AM = -0.010 ± 0.028; Entropy = -0.038 ± 0.04) diminishes singing induced variability relative to that seen in GFP-expressing controls (green bars; n = 9 syllables; Duration = -0.128 ± 0.071; AM = -0.065 ± 0.035; Entropy = -0.091 ± 0.034) In contrast, overexpression of FoxP2.10+ (red bars; n = 13 syllables; Duration = 0.070 ± 0.054; AM = 0.088 ± 0.047; Entropy = 0.048 ± 0.029) causes singing to induce a relative state of invariability. Bar heights represent the average effect size for all syllables within the virus construct group ± SEM. * denotes significant result in one-way ANOVA (Duration: $F(2,35) = 3.95$, $p = 0.028$; AM: $F(2,35) = 3.96$, $p = 0.028$; Entropy: $F(2,35) = 3.63$, $p = 0.037$) and Tukey's HSD post-hoc test ($p < 0.05$). **d)** Learning curves plot the relationship between percentage similarity to tutor as a function of time. Animals overexpressing GFP (green; letter 'A'; n = 7 birds; ~65d similarity = 67.2 ± 6.64%) or FoxP2.10+ (red, letter 'B'; n = 5 birds; ~65d similarity = 75.8 ± 2%) learn significantly better than those overexpressing FoxP2.FL (grey, letter 'C'; n = 5 birds; ~65d similarity = 44.3 ± 10.1%). Data are binned by day (top panel; bold points represent group mean and shifted smaller points are individual birds) or by individuals (bottom panel). Statistically significantly different groups tested by one-way ANOVA (Bin 1: ~40d $F(2,11) = 6.06$, $p = 0.016$; Bin 3: ~55d $F(2,13) = 6.04$, $p = 0.014$; Bin 4: ~60d $F(2,14) = 9.94$, $p = 0.002$; Bin 5: ~65d $F(2,14) = 4.76$, $p = 0.026$) and Tukey HSD post-hoc test ($p < 0.05$) are denoted by capital and lowercase lettering. **e)** Exemplar motifs of a tutor and three of his 65d pupils, each of

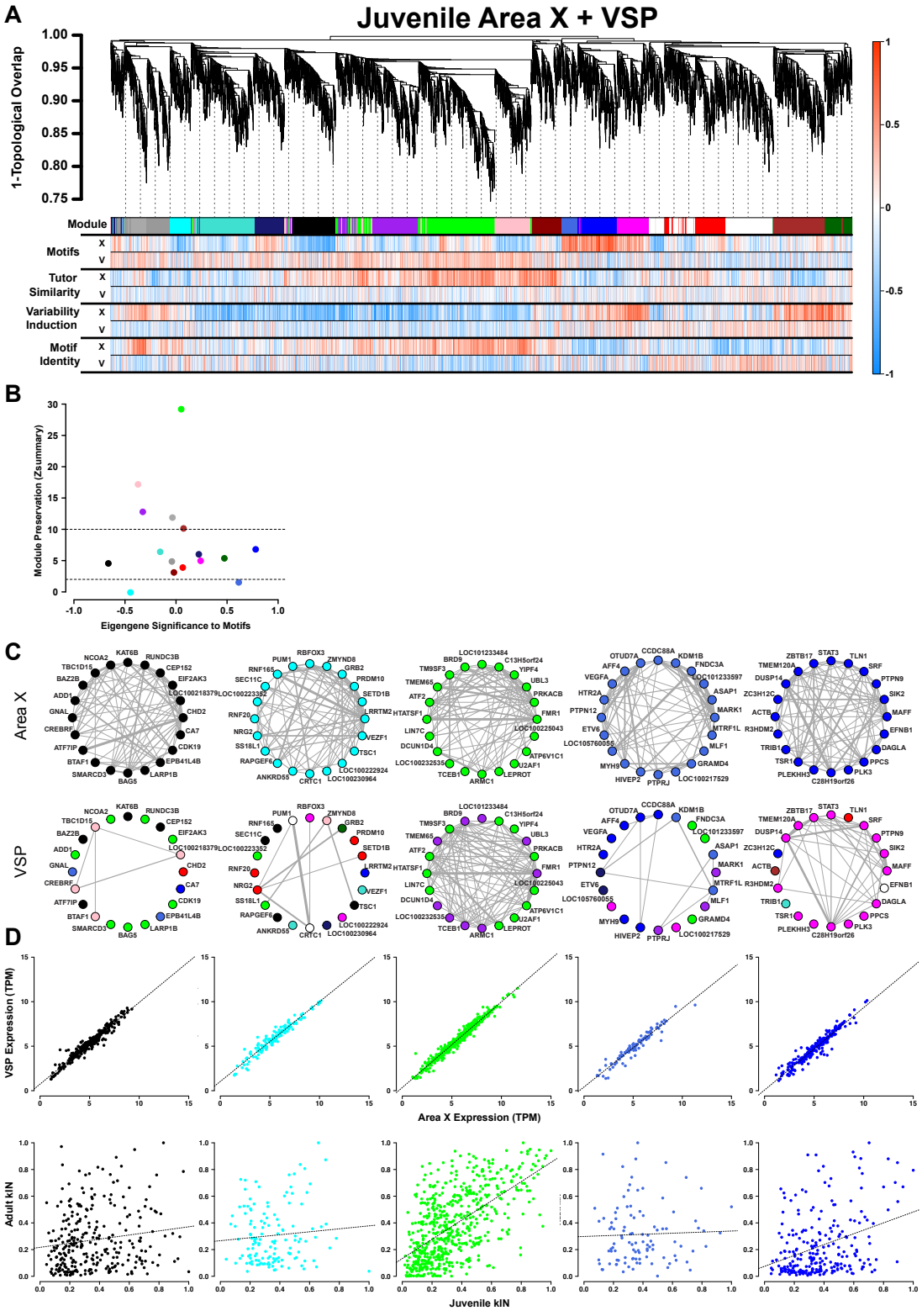
which was injected with a different viral construct at 30d. These examples illustrate the percent similarity depicted in panel D. **f)** Summary of the learning and variability phenotypes observed after virus injection.

Figure 2-3: WGCNA yields behaviorally relevant modules



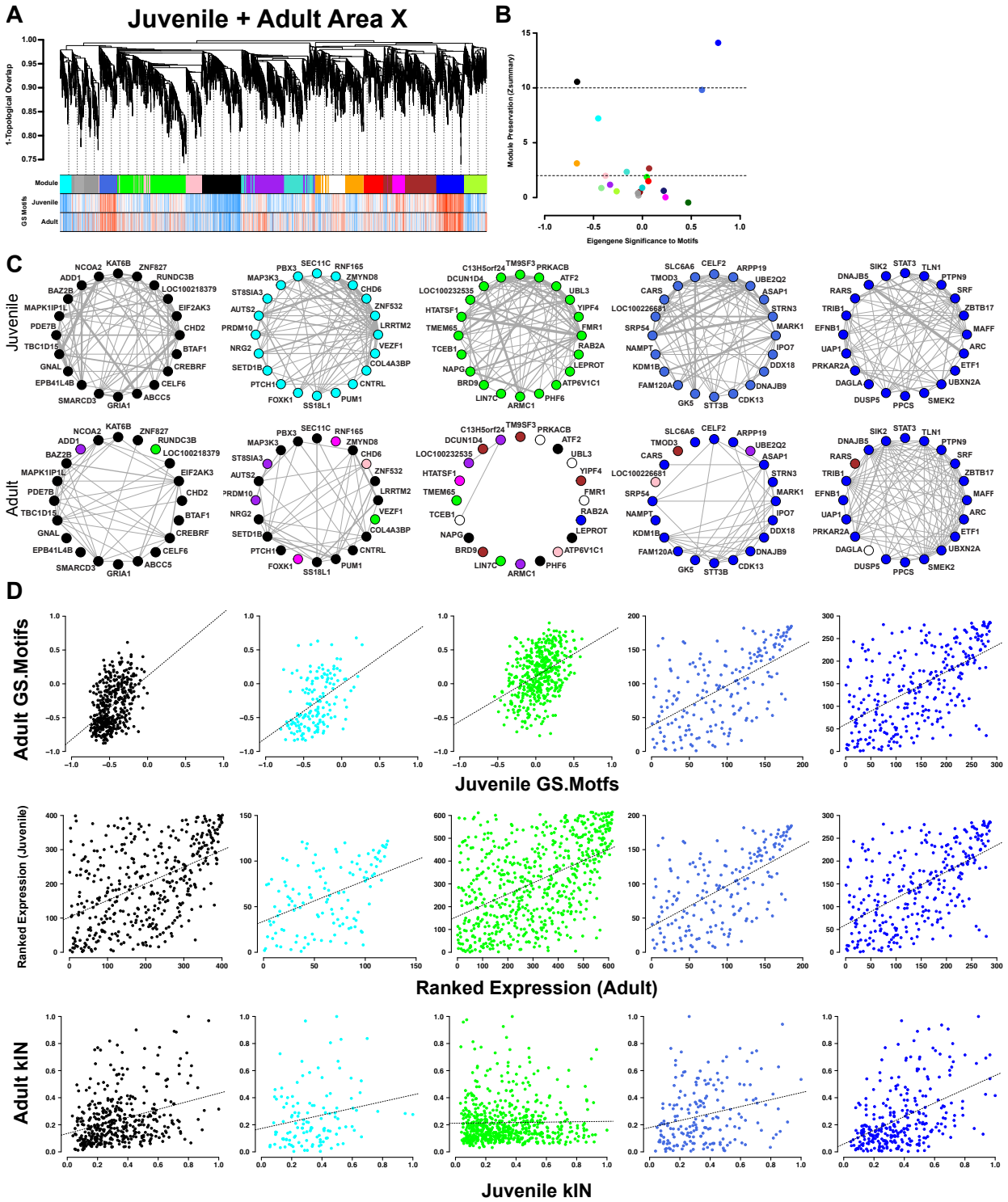
a) Dendrogram (top) illustrates the topological overlap between genes in the juvenile Area X network. Modules delineated by automated tree trimming are shown below and are depicted by arbitrary colors. Beneath the color bar, gene significances to the quantified behaviors (number of motifs sung, tutor similarity, acute variability changes, and overall variability; see Results) are indicated by a heatmap wherein red indicates a positive correlation and blue indicates a negative correlation (see B for scale). **b)** Correlations between module eigengenes and each behavior are presented as a heatmap. The Pearson's ρ and, in parentheses, Student's asymptotic p-values for modules where $p \leq 0.05$ are displayed. **c)** For all significant module-trait correlations, the relationship between gene significance and module membership is plotted for each gene in the module. Dashed lines represent the linear regression and the Pearson's ρ ("cor") and p-value as determined by Fisher's z-transformation are indicated above each plot. **d)** The average whole network connectivity (kTotal) within each module reveals that the purple, green, and pink modules are composed of the most strongly connected genes in the network. **e)** Term significances for the black, darkred, and green modules for are indicated for disease, gene ontology biological process and molecular function, as well as for pathways for categories annotated as 'neuronal' in the GeneCards GeneAnalytics software. **f)** Network plots of the modules presented in panel E where nodes represent genes scaled by the node's intramodular connectivity and edge width displays the topological overlap between genes.

Figure 2-4: Area X singing related gene coexpression patterns are not preserved in VSP



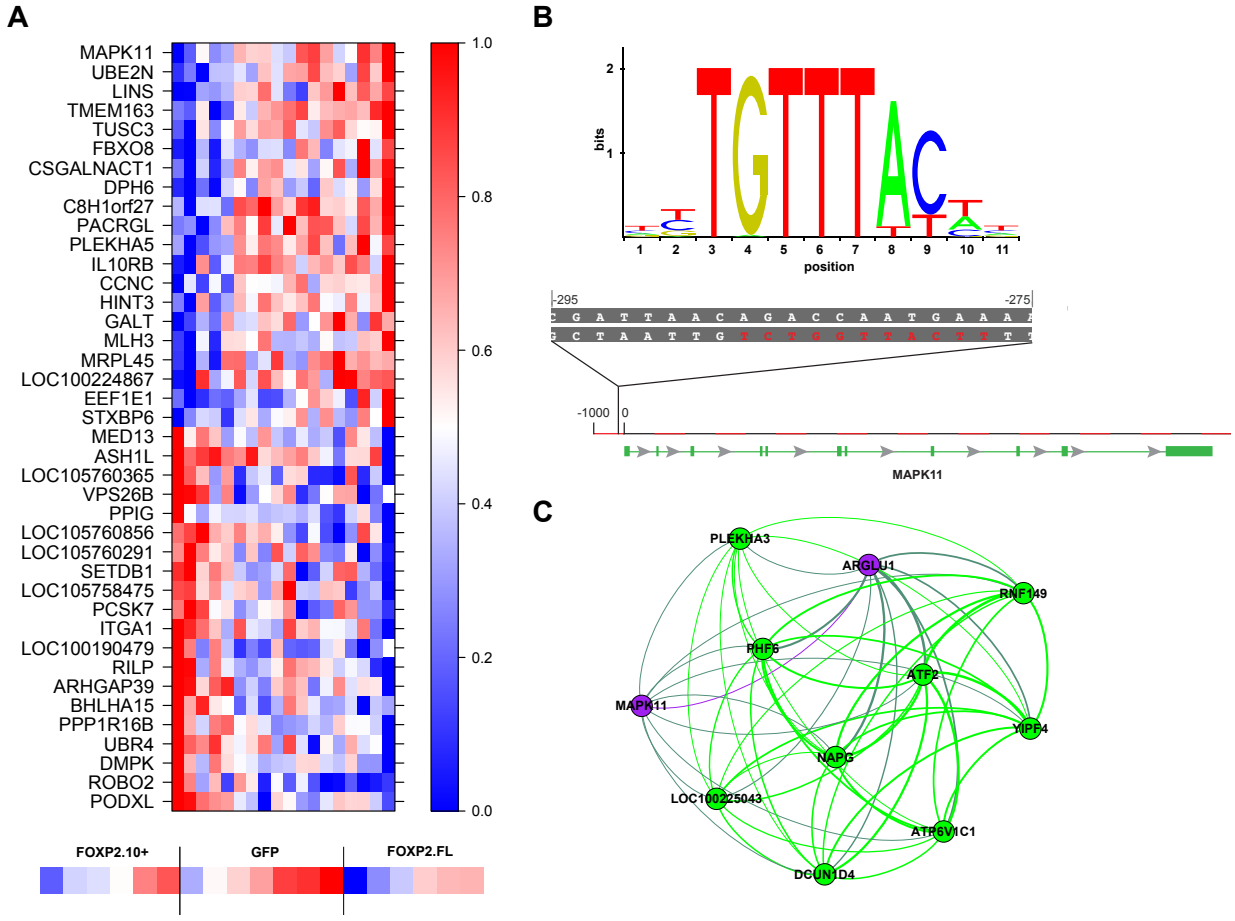
a) Dendrogram (top) displays the topological overlap in Area X between genes common to both Area X and VSP networks. Beneath, the module assignments and the gene significances for each gene as calculated using expression from VSP ("V") or Area X ("X") for all behaviors are quantified as in Fig. 3A. Module colors are consistent with those presented in Figure 3. **b)** Module preservation (Zsummary) for all modules that were present in both Area X and VSP displayed as a function of ME correlation to motifs. Lower and upper dashed horizontal lines indicate thresholds for low and high preservation, respectively. **c)** Circle plots display the adjacencies between the 20 most well-connected genes in the Area X black, cyan, green, royalblue, and blue modules. The adjacency between genes is indicated by edge thickness. Genes grouped together in the black, cyan, royalblue, and blue song modules in Area X have numerous and strong connections. Those connections are weakened or nonexistent in VSP such that genes sort into different modules in VSP. In contrast, the green learning module genes maintain their common grouping and connections in VSP. **d)** Raw gene expression is tightly correlated between Area X and VSP for the genes in the black, cyan, green, royalblue, and blue modules (top). Only the intramodular connectivity of the genes in the green module is correlated between Area X and VSP (bottom). Dashed lines represent the linear regression.

Figure 2-5: Area X song but not learning modules are preserved into adulthood



a) Dendrogram (top) displays the topological overlap in juvenile Area X between genes common to both juvenile and adult Area X networks. The module assignments and the gene significances to motifs in juveniles and adults are presented below. Module colors are consistent with those presented in Figure 3. **b)** Module preservation (Zsummary) for all modules that were present in both juvenile and adult Area X displayed as a function of ME correlation to motifs. Lower and upper dashed horizontal lines indicate thresholds for low and high preservation, respectively. **c)** Circle plots display the adjacencies between the 20 most well-connected genes in the juvenile Area X black, cyan, green, royalblue, and blue modules. The adjacency between genes are indicated by edge thickness. Genes grouped together in the black, cyan, royalblue, and blue song modules in Area X have numerous and strong connections that are mostly maintained in adulthood. The densely interconnected green learning module genes found in juveniles do not maintain these relationships in adulthood. **d)** Strong positive correlations between gene significance to motifs exist for all modules (top row). Ranked expression values for the genes in each module also show positive correlation (middle row). Intramodular connectivity is more positively correlated between ages for the black, cyan, royalblue, and blue song modules than for the green learning module.

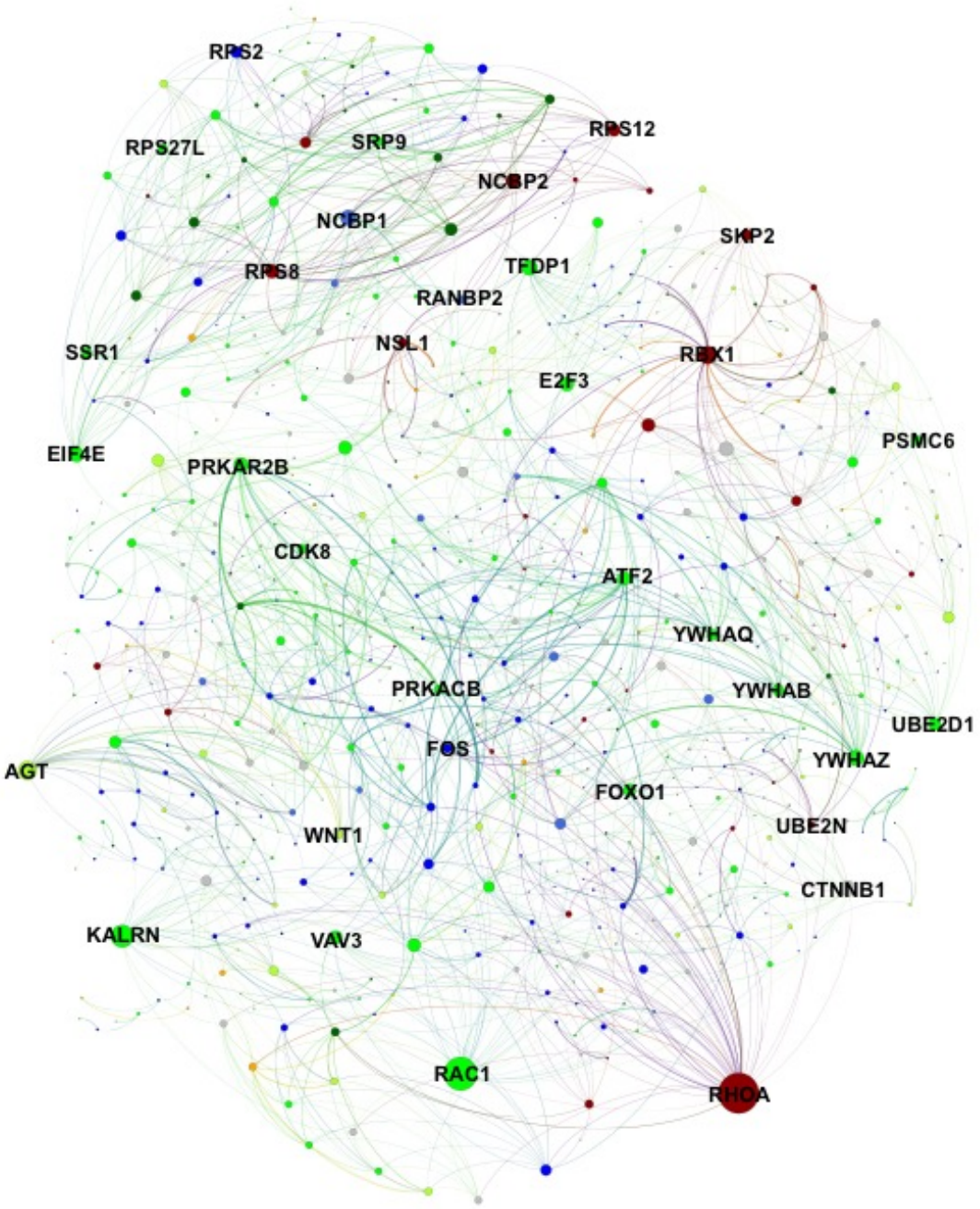
Figure 2-6: Gene significance and network position implicate MAPK11 as a molecular entry point to vocal learning mechanisms



a) The 20 genes with the highest to lowest gene significances to tutor similarity (sorted from top to bottom) are shown. Each column represents a bird and columns are sorted in order of increasing tutor similarity from left to right. Gene expression is scaled such the highest and lowest expression across samples have the brightest shade of red or blue, respectively. Below, expression of MAPK11 is shown again, here separated by virus group and then sorted by increasing tutor percentage similarity. **b)** The FoxP2 binding sequence as annotated by the JASPAR database (top) and a potential binding site found in the promoter upstream of MAPK11. **c)** MAPK11 and its 10

closest network neighbors, including green learning module members and hub gene ATF2, as defined by topological overlap.

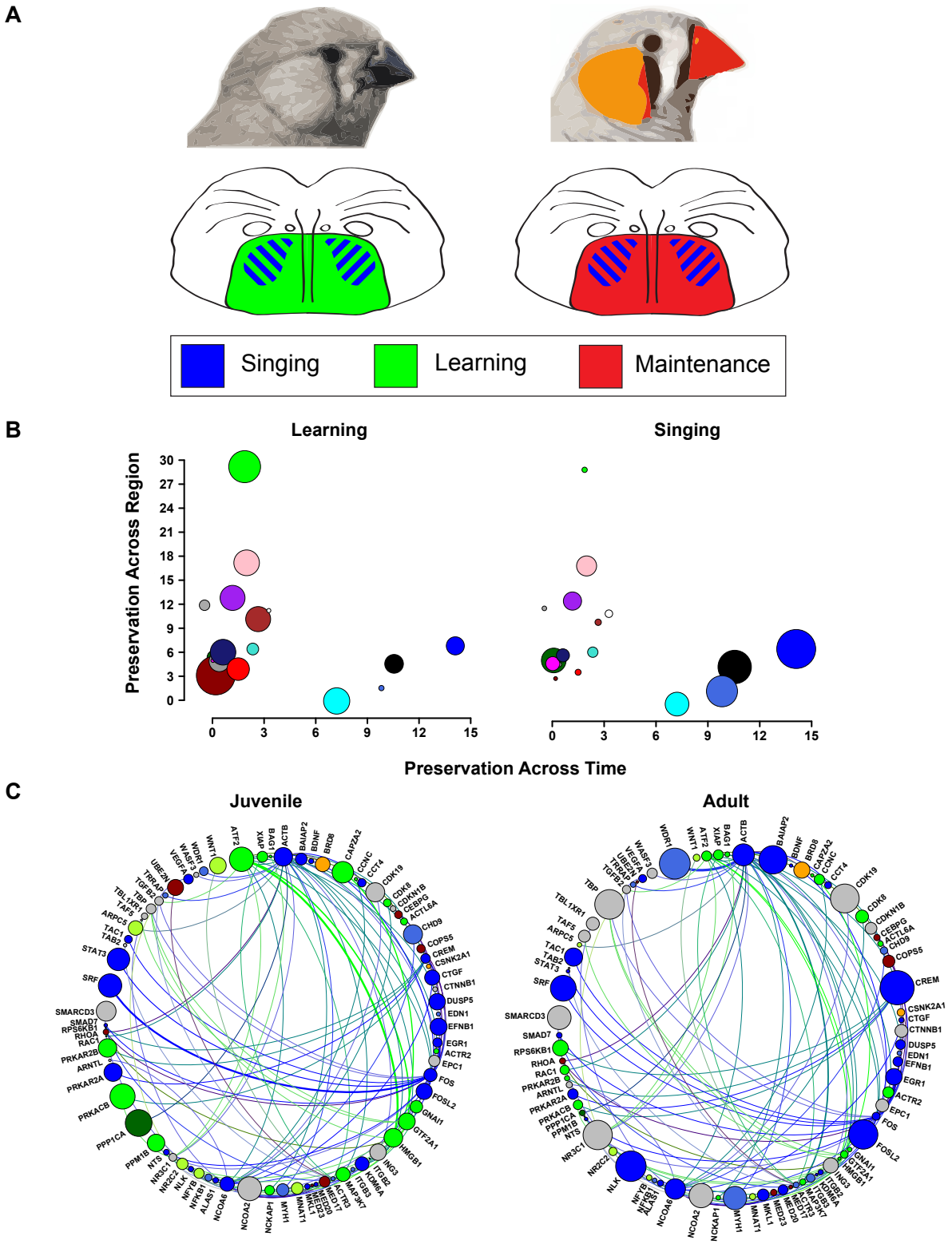
Figure 2-7: Protein-level interactions between song and learning module genes in juvenile Area X



A protein interaction network plot using the STRING database between genes in learning (darkred, green, greenyellow) and song (black, blue, darkgreen, orange, royalblue) modules. Nodes are

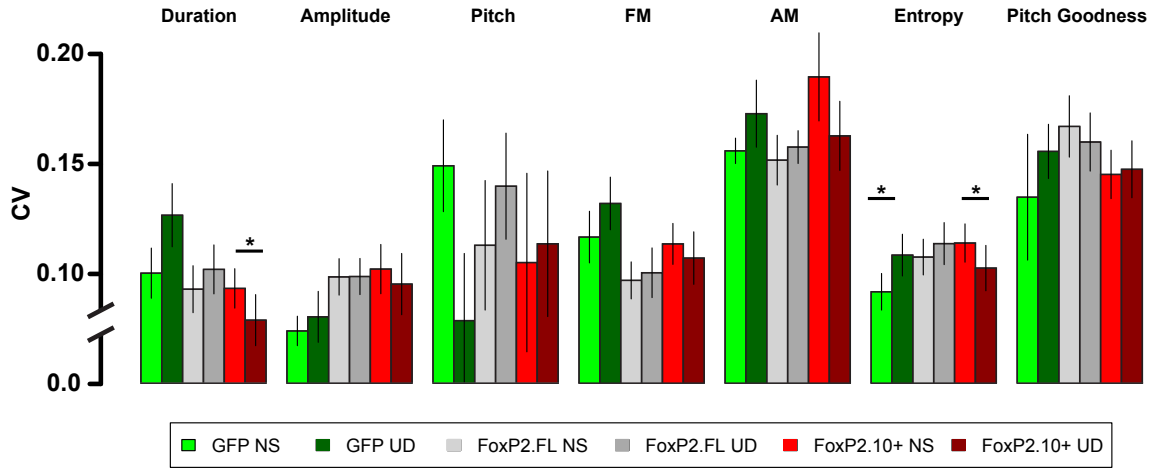
scaled by number of connections. Edge width is determined by scaling the STRING protein interaction confidence score for the two nodes by the product of each node's intramodular connectivity. Interactions within learning or song modules are omitted for clarity.

Figure 2-8: Changes in vocal plasticity state between juvenile and adult birds



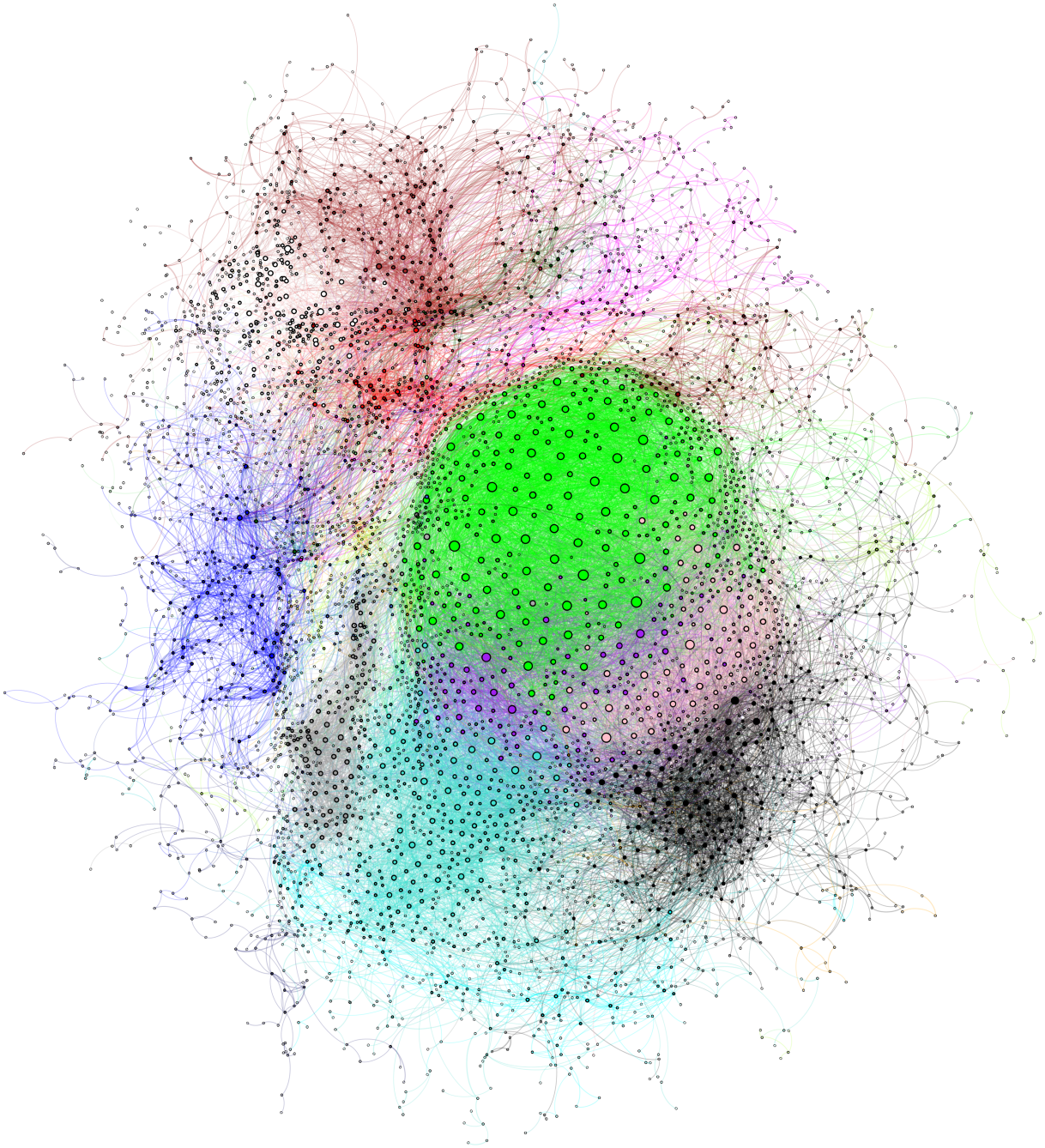
a) The juvenile straitopallidum (left) exists in a plastic state in which genes in learning modules (e.g. green) are densely interconnected and of high importance in the network. Simultaneously, singing driven gene coexpression patterns occur. In the adult striatopallidum (right), song modules exist as they do in juveniles, but the learning modules do not. **b)** Area X modules in the juvenile brain are plotted to emphasize their preservation in adult Area X (x axis) and juvenile VSP (y axis). Points representing the module colors are scaled by the module's absolute correlation to learning (left) or the absolute correlation to singing (right), emphasizing the preservation of singing coexpression patterns into adulthood and learning coexpression patterns in the juvenile striatopallidum. **c)** Genes in song or learning modules that are within two steps of ATF2 in the high-confidence protein interaction network are shown. Nodes are scaled by intramodular connectivity in juveniles (top) or adults (bottom) with edge width indicative of adjacency between genes in the coexpression network. The change in coexpression patterns across age groups causes decreased connectivity of many learning-related genes, driving an alteration in the network's landscape which may underlie the transition from song learning to song maintenance.

Figure 2-S1: Raw acoustic feature variability in the NS and UD conditions by virus group



The raw acoustic feature CVs transformed by the calculation in Figure 1D show the variability relationship between NS and UD contexts for all measured acoustic features. For most song features, UD singing drives increases in CV in the GFP group. This effect is blocked or reversed in the FoxP2.FL and FoxP2.10+ groups, respectively. Notably, the songs of FoxP2.10+ animals following 2 hours of UD song were significantly less variable than those after 2 hours of silence. Asterisks indicate a significant difference ($p < 0.05$) in a paired resampling test within virus construct.

Figure 2-S2: Juvenile Area X gene coexpression network



The gene coexpression network is displayed with genes represented as nodes and colors indicating the module assignment of each gene. Nodes are scaled by their degree and edge color is the

combination of the module colors of nodes connected by the edge. Poorly connected nodes are excluded.

Chapter 3: VoICE: A Semi-Automated Pipeline for Standardizing Vocal Analysis Across Models

Zachary D. Burkett, Nancy F. Day, Olga Peñagarikano, Daniel H. Geschwind, and Stephanie A. White

Abstract

The study of vocal communication in animal models provides key insight to the neurogenetic basis for speech and communication disorders. Current methods for vocal analysis suffer from a lack of standardization, creating ambiguity in cross-laboratory and cross-species comparisons. Here, we present VoICE (Vocal Inventory Clustering Engine), an approach to grouping vocal elements by creating a high dimensionality dataset through scoring spectral similarity between all vocalizations within a recording session. This dataset is then subjected to hierarchical clustering, generating a dendrogram that is pruned into meaningful vocalization "types" by an automated algorithm. When applied to birdsong, a key model for vocal learning, VoICE captures the known deterioration in acoustic properties that follows deafening, including altered sequencing. In a mammalian neurodevelopmental model, we uncover a reduced vocal repertoire of mice lacking the autism susceptibility gene, *Cntnap2*. VoICE will be useful to the scientific community as it can standardize vocalization analyses across species and laboratories.

Introduction

Though no animal model adequately captures the sophistication of language, ethological study of vocal communication has yielded valuable insight to its evolution and physiological basis. The learned songs of oscine songbirds are well-studied in the laboratory environment. The discrete brain circuitry, shared molecular reliance with humans, requirement of auditory feedback for maintenance, and parallel anatomical loops for producing learned vocalizations have made songbirds a powerful model for speech and language[55,105]. A key strength of rodent model systems is their genetic tractability, allowing researchers to precisely manipulate potential disease genes or neural circuits. In contrast to birdsongs, the ultrasonic vocalizations (USVs) generated by rodents are largely innate yet none-the-less provide an important phenotypic dimension[106,107]. As interest in a comprehensive analysis of social communication signals increases, the need for standardization across models becomes apparent. To meet this challenge, we designed an analysis pipeline into which any type of discrete vocal element (VE) can be input, and the output of which provides valid results in both acoustic and syntactical (defined here as the sequence in which vocal elements occur) domains.

The learned courtship song of male zebra finches (*Taeniopygia guttata*) exists in the audible range (20 Hz to 20 kHz), and is hierarchically composed of notes, syllables, motifs, and bouts[108]. Notes, the smallest unit, are defined as a region of a syllable that maintains a temporally continuous frequency pattern. Syllables are discrete vocal units that contain one or more notes. Motifs are repeated units of 3 or more syllables lasting ~1 second. A bout of song is composed of multiple motifs, frequently with repeated introductory syllables at the beginning of

each bout. Analytical methods for quantifying the acoustic and syntactical components of song[68,109-112] have proven efficacious, but their cross-species application is limited. Moreover, they exhibit a combination of drawbacks, including low resolution in quantifying acoustic differences between syllables due to calculating distance in a limited acoustic feature space, and/or a low throughput due to extensive user parsing of the data. Lastly, methods for quantifying syntax require the user to explicitly dictate the number of syllable types that emerge from the analysis or to incorporate the acoustic domain into the calculation of sequential similarity.

Mouse (*Mus musculus*) USVs occupy a frequency range that extends above 20 kHz and are generated in a variety of contexts. For example, when neonates become separated from the dam, they emit retrieval calls. Adult male mice make courtship calls when presented with females or their pheromones, whereas females produce calls when searching for their pups or in the presence of another female[113]. USVs are mostly narrowband, with spectral power distributed across a limited frequency range. In contrast to the stereotyped adult songs of zebra finches, a high degree of call-to-call variability exists in the mouse ultrasonic repertoire. Despite this variability, 10 distinct retrieval call categories have been defined and adopted (or modified), allowing for quantitative analyses of vocal signals generated by neonatal mice[114]. The process of individually parsing these calls into one of the categories, however, is labor-intensive, requiring manual inspection of each call and a subjective assignment into a category. This strategy leads to poor reproducibility of classifications, reduces the throughput of the analysis, and provides numerous opportunities for error. Efforts to automate this procedure resulted in the empirical derivation of a considerably reduced number of distinct call types, whose broad adoption would require a

complete shift in how researchers consider the extent of the mouse vocal repertoire in order to maintain comparable and reproducible results across laboratories[115].

To circumvent the limitations of existing vocal analysis techniques in both songbirds and rodents, we developed VoICE, software that utilizes acoustic similarity relationships between VEs to generate high dimensional similarity matrices, which are then input to a hierarchical clustering algorithm. We glean distinct syllable or call "types" by using an algorithm that automatically prunes clusters from hierarchical trees[45]. MATLAB is utilized for the user interface and similarity scoring, while scripts written in the R statistical programming language perform clustering and dendrogram trimming. We validate VoICE by applying it to vocalizations from adult male zebra finches and to mouse pups, and then compare the results to those obtained through manual parsing of finch syllables or mouse calls. We then illustrate the utility of VoICE by quantifying the phonological (i.e. acoustic) and syntactical consequences of deafening in the zebra finch. Further, use of VoICE replicates the finding of reduced numbers of retrieval calls in pups lacking the *Cntnap2* gene, an established model of autism[116], and also uncovers changes in the repertoire of these animals. These findings establish this approach as a reliable, high-throughput method that faithfully captures known features of avian and rodent vocalizations and is capable of uncovering novel changes in this critical phenotypic trait.

Methods

Software

All scripts and a manual containing step-by-step instructions for installation and conducting analyses are available at https://www.ibp.ucla.edu/research/white/CODE.html.

Subjects

Finch: Three adult (>125 days) zebra finches (*Taeniopygia guttata)* were removed from our breeding colony (13:11 hour light/dark cycle). All animal husbandry and experimental procedures were in accordance with NIH guidelines for experiments involving vertebrate animals and approved by the University of California, Los Angeles Institutional Animal Care and Use Committee. Birds were fed seed and calcium-enriched (Calciboost, The Birdcare Company, Gloucestershire, UK) water ad libitum, provided with weekly nutritional and environmental supplements (hard-boiled chicken egg, fresh carrots and komatsuma, millet sprays, bathing water).

Mouse: All animal husbandry and experimental procedures were in accordance with NIH guidelines for experiments involving vertebrate animals and approved by the University of California, Los Angeles Institutional Animal Care and Use Committee. Pups (P7) were removed from the dam and placed in individual soundproof chambers equipped to record ultrasonic vocalizations for 5 minutes using an ultrasonic microphone with a flat frequency response up to 150 kHz and a working frequency response range of 10-180 kHz (CM16, Avisoft, Germany). To avoid any potential confounding effects due to temperature, the room was maintained at 21°C.

After recording, pups were tattooed and tail tissue was obtained to perform genotyping before returning them to the dam. Recordings were performed in the UCLA behavioral testing core. Mice were kept in 12 hr light/12 hr dark cycle and had ad-lib access to food and water. All procedures were performed in accordance with the UCLA Institutional Animal Care and Use Committee.

Finch deafening surgeries

Following 2-3 weeks of baseline recording, two finches were deafened by bilateral removal of the cochlea as described by Konishi[117]. Briefly, birds were anesthetized with inhalant isoflurane and secured on a rotary table. Under a dissection microscope (OPMI pico, Carl Zeiss Meditec, Inc., Dublin, CA), a small area of skin as well as the tympanic membrane overlaying the middle ear cavity was removed using iridectomy scissors, followed by the removal of the columella, allowing visualization of the cochlea. A small hook made of tungsten fiber was used to extract the cochlea. Removal of an unbroken cochlea indicated the initial success of the surgery, which was later confirmed by deterioration of song. One additional bird underwent sham operations to control for any potential effects of the surgical procedure itself. Sham operations consisted of the anesthetic protocol and skin removal as the deafened birds, but without damage to the tympanic membrane or removal of the columella or cochlea. Following the procedure, Neosporin was applied to each ear and the animal was monitored in its recording chamber to ensure that the birds did not show life-threatening vestibular damage.

Recording and sound preprocessing

Finch: Songs were recorded when birds were singly housed in sound attenuation chambers (Acoustic Systems; Austin, TX). Songs were recorded biweekly for approximately 7 months using Sound Analysis Pro[68] (v. 2011.107) recorder using Countryman EMW Omindirectional Lavalier microphones (Countryman Associates; Menlo Park, CA) attached to the chamber ceiling at a fixed location. Sounds were digitized using a PreSonus Firepod at a sampling rate of 44.1 kHz and 16 bit depth. For each day of analysis, the first ~300 syllables were semi-automatically extracted using the SAP "Explore & Score" module. Raw sound recordings were opened in SAP's "Explore & Score" module (Figure 3-1), then a segmentation threshold was applied to delineate syllable boundaries. Syllable tables were then built by highlighting syllables sequentially and exporting their records to an SQL table, where individual rows represent syllables in the sequence in which they were sung, adjusting segmentation parameters as needed to ensure optimal syllable start and end boundaries. The syllable tables were then imported to R, where the syllable onset and duration data were used to automatically clip each syllable from the raw recording to its own WAV file.

Mouse: Calls were recorded over a single 5-minute recording session for each mouse at postnatal day 7 (P7) between 9:00 AM and 10:00 AM during their light cycle. To avoid any potential confounding effects due to temperature, the room was maintained at 21ºC. To isolate individual calls from the raw recordings, we used a MATLAB script that calculated the amplitude envelope in each file and clipped into individual WAV files all sounds that passed the threshold for longer than 2 msec. Calls were considered complete after the clip fell below threshold for 10 msec. Finally, all sounds longer than 150 msec were considered to be noise and discarded from the analysis. To provide the highest quality input data to our clustering workflow, we then manually inspected WAV files and discarded those that contained only noise.

Similarity scoring

Finch: Similarity scores between syllables were calculated using a MATLAB algorithm adapted from the C++ code for the SAP similarity batch function. Briefly, syllables were quantified millisecond-by-millisecond based on four features: Weiner entropy, frequency, FM, and pitch goodness (a measure of sound periodicity). These features were calculated in 9.3ms bins with sliding 1-ms steps with a subprogram of SAP, ported to MATLAB by S. Saar. Upon completion of the similarity batch, global similarity (GS) scores (Figure 3-2) for each syllable-syllable comparison were calculated by dividing the product of similarity, accuracy, and sequential match ("temporal overlap") by 10,000, a metric employed previously to describe the overall likeness between song types following deafening[118].

Mouse: Our original intent was to score acoustic similarity between USVs in a fashion similar to bird song syllables. Due to the great difference in bandwidth between bird song syllables and mouse USVs, this metric indicated a high level of similarity between all calls. Therefore, we designed a different metric focused on describing relationships between the frequency contours of the mostly narrowband calls. All recordings were sampled at 250 kHz, then filtered so that only spectral data between 40 and 120 kHz was considered. Each call was transformed to a spectral derivative in MATLAB. Calls were then binned into windows of 44 samples each and the average pitch across all samples was calculated for each window. The median pitch across every 5 consecutive windows was then calculated and used, therefore allowing us to describe the overall pitch for every ~0.9 msec of each call.

Once frequency contours were calculated, the pitch scores were then compared between all calls in a pairwise fashion. First, the Pearson correlation of raw pitch was calculated for all overlapping windows between each pair of calls in order to describe the similarity of frequency contours. The correlation of raw pitch is independent of time and the frequency range in which the two calls reside, therefore scoring similarity simply by correlation in pitch could lead to similar frequency contours at vastly different frequency ranges appearing highly alike. Therefore, we took additional measures to weight pitch correlation to more accurately describe the relationships between calls. First, the absolute difference in pitch at each overlapping ~0.9 msec window was calculated and averaged across all windows, divided by the maximum possible difference in pitch (or 80 kHz, as dictated by the frequency range over which we analyzed spectral data), then subtracted from 1. Calls occupying a similar frequency range will have very low absolute differences in pitch, leading to a pitch difference score close to 1. Finally, to account for differences in duration between pairs of calls, the temporal overlap was calculated. The product of these three measures, pitch correlation, scaled pitch difference, and temporal overlap resides on a -1 to 1 scale and was used as the input to the hierarchical clustering function (Figure 3-3).

Both species: Before implementation of the hierarchical clustering function, acoustic similarity scores were transformed to an M x M (M=number of vocalizations) similarity matrix. For finch syllables, the Euclidean distance between all syllable-syllable pairs in the GS matrix were then calculated, generating an M x M distance matrix, which is used as the input to an average linkage hierarchical clustering algorithm. For mice, all values in the weighted correlation matrix are subtracted from 1 and used as the dissimilarity input to the clustering algorithm.

Hierarchical clustering

Both species: The dissimilarity matrix generated in the previous step is used as input to the flashClust function[119] in the weighted gene co-expression network analysis (WGCNA;[67]) package in R (http://r-project.org). Syllables are clustered by average linkage hierarchical clustering to generate a dendrogram. Hierarchical clustering is desirable for this type of analysis, as it does not require dictation of a number of call types as other clustering methodologies do. The hierarchical cluster tree is trimmed using the Dynamic Tree Cut R package 'Dynamic Hybrid' tree cut function[45]. As opposed to dictating a static tree cut height and generating clusters based on which branches remain together following trimming at this user-dictated height, this algorithm climbs bottom-up through hierarchical trees and derives clusters in an automated fashion based on user-dictated parameters. Despite its mostly automated function, the tree-cutting algorithm is tunable to an extent that can influence the number of clusters that are gleaned from the hierarchical tree. To be as rigorous as possible, we dictated that the minimum cluster size to be a single vocalization and set the deepSplit parameter to its highest level of sensitivity in cluster detection. We offer the user the ability to tune these parameters to suit their own needs.

Eigensyllable/eigencall calculation

Both species: Following tree-trimming, each multi-vocalization cluster is given a unique color name and is described by calculating a cluster "eigensyllable" or "eigencall", defined as the first principal component of the acoustic similarity scores within the cluster as determined by singular value decomposition[67]. The eigensyllable or eigencall can be considered as a single representative of the acoustic properties of all the vocalizations within a cluster.

Iterative cluster merging

Finch: Since it is possible to ascertain a "correct" number of distinct syllable types in zebra finch song, small clusters initially derived through the divisive trimming of the dendrogram are merged together. The Pearson's product-moment correlation between each cluster eigensyllable is calculated and clusters whose eigensyllables correlate above a user-provided Pearson's rho threshold are merged together. This process then repeats until no cluster eigensyllables correlate above the given threshold. Since, ultimately, the number of cluster merges performed by the automated tree-trimming algorithm determines the number of clusters and, therefore, the number of syllable types in the animal's repertoire, we sought to develop an iterative procedure to empirically derive the ideal merge threshold.

The iterative procedure is as follows: the hierarchical tree is created, then the automated tree-trimming algorithm is applied to the tree using the most divisive parameters possible, creating numerous small clusters. Each small cluster is represented by an eigensyllable. The iterative procedure then begins, where clusters whose eigensyllables correlate at or above 0.99 are merged. The rho is then decreased to 0.98 and the process repeats. This process repeats at each rho until reaching 0. At each step, the number of clusters and the average intracluster global similarity score (IGS) are calculated. As the merging threshold approaches zero, the number of clusters and the IGS for each merged pair of clusters decreases. One would expect the ideal cluster number (n) to remain stable over a large range of merging thresholds, as this would indicate that each cluster is sufficiently dissimilar to every other cluster that it will not be merged (unless the threshold for merging becomes so low as to allow for enough intracluster variability to override concrete

differences in syllable spectral properties). When the merging threshold decreases to a point where all clusters that do not display a level of anticorrelation are joined together, a floor for the possible number of syllable types is reached. Upon completion of the iterative cluster merging, the IGS and the number of syllables in each cluster at each cluster n that remained stable over multiple merging steps are returned to the user. The user then selects the appropriate cluster n by considering the degree of IGS and the expected number of syllable types in the animal's repertoire.

Once an appropriate merging threshold has been selected, a list is returned to the user containing: (1) the data, in the form of an M x M (m = number of syllables) distance matrix, worked on by the clustering algorithm, (2) the song syntax before applying any merging techniques, (3) the song syntax at the derived merging threshold, (4) the eigensyllables, and (5) the proportion of variance in each cluster explained by its eigensyllable. WAV files of every syllable within each cluster are then generated and presented to the user for purposes of error checking. Manual error correction is accomplished by changing the cluster ID for each syllable within the merged song syntax component of the list generated during clustering. Once clusters are finalized, the syntax is returned and the acoustic data are sorted into tables for each cluster.

USV cluster quality control and eigencall-based classification

Unlike the crystallized and highly stereotyped vocalizations of adult finches, a high degree of call-to-call variability in the mouse ultrasonic vocal repertoire exists. Applying a merging strategy similar to the one implemented for finches resulted in few cluster ns that stayed stable over multiple merging thresholds, indicating that the number of distinct call types is highly sensitive to the amount of variability allowed within the cluster, making it impossible to ascertain

the “correct” number of syllable types. This drove us to implement a cluster quality control step before dictating all calls within a cluster as “the same.” For each cluster defined by the tree-cutting algorithm, the Pearson correlation between all calls within the cluster and the cluster eigencall are calculated. Clusters whose average Pearson correlation is below a user-defined threshold (for all data described in this manuscript, a correlation threshold of ≥ 0.8 was used) are considered insufficiently cohesive and syllables within are classified manually. Clusters that do show sufficient correlation are then represented by the single syllable within them that displays the highest Pearson correlation to the cluster eigencall.

Following quality control, individual calls from dissolved clusters and/or single-call representatives of individual clusters were classified into the canonical types as described by Scattoni et al.[114]. In some cases, our method for automatically clipping calls from the raw recordings resulted in multiple calls being clipped into one WAV file. We therefore added categories called “double” and “triple” to reflect when this occurred. Finally, we observed a small but not unsubstantial number of calls that did not display the characteristics of any of the canonical calls. Instead of forcing these into a category in which they do not fit, we added a “miscellaneous” category to which we assigned these syllables.

Assignment of syllables to established clusters

The preceding clustering step generates syllable clusters for a single animal within one recording session. In order to relate a second recording session to clusters created in an initial session, a similarity batch is first performed between all syllables in Session B and a subset of syllables in Session A (sorted by cluster). GS is calculated for all comparisons. A subset of

syllables is selected from each Session A cluster to decrease computation time in the similarity batch. Should every member of each Session A cluster be included, nearly redundant similarity comparisons between a Session B syllable and highly alike Session A syllables would occur, increasing computation time without providing useful similarity data. The syllable subset from each session one cluster is determined by calculating the correlation between all syllables in each cluster to their respective cluster eigensyllable. The correlations are then ranked and the top 10% are used as representatives of the cluster, as these syllables statistically represent the greatest proportion of variance within the cluster. In our initial testing, using the top 10% resulted in nearly identical syllable assignment results as using the entirety of each cluster with a ~90% savings in processing time, though the user ultimately can select the percentage of the cluster to use. To prevent inappropriate syllable assignment to a cluster in the case of an improvised syllable during the second recording session, a GS floor is dictated by the user, below which a syllable will not be assigned to a cluster. The assignment of the GS floor is at the discretion of the experimenter, with a lower floor used in cases where acoustic similarity between recording sessions is not expected to be high, such as in the case of assigning a juvenile pupil bird's syllables to a tutor's clusters or syllables obtained following a deafening procedure to clusters created from pre-deafening song.

Syllables from the second recording session are then considered one at a time in their relationships to the representatives of the established clusters. For all established clusters of which an unassigned syllable shares an above-threshold GS, a one-way ANOVA and pairwise comparisons post-hoc test are performed to determine whether the unassigned syllable shows a statistically significant relationship after Bonferroni p-value correction with one cluster above all others. In the case that a significant relationship is determined, the unassigned syllable is given the

same cluster ID as the established cluster. In the case that no significant relationship is observed, yet multiple clusters existed above the GS threshold, the unassigned syllable is passed into a tiebreaking queue (see below). Lastly, in the occurrence of an unassigned syllable that does not show an above-threshold global similarity relationship with any established cluster, it is deemed a novel syllable and passed into a queue for later derivation of the number of novel syllable types present.

Once the first pass through syllable-to-cluster assignment has finished, a round of tiebreaking occurs. All syllables that were passed into the tiebreaking queue have met the following conditions: (1) showed an above-threshold GS with more than one established cluster and (2) did not show a statistically significant relationship with one cluster above all others. From this point, the user is allowed to view the spectrograms of the unassigned syllable and the syllable within each established cluster that displays the highest correlation with that cluster's eigensyllable. The user then clicks a button to assign the syllable to any cluster or add it to the unassigned syllable queue.

Upon completion of the previous two cluster assignment steps, the remaining syllables that did not get passed into a cluster are considered in an effort to determine the number of syllable types present in the second recording session that are not present in the first (e.g. syllables sung by a pupil that are not in its tutor's vocal repertoire). A similar procedure to the initial clustering step is performed, utilizing the GS scores between each unassigned syllable and each representative of each session one cluster. Unassigned syllables that show similar relationships to the established clusters are likely similar to one another. Using this logic, first, an M x N (M = number of

unassigned syllables from session two, N = total number of cluster representatives from session one) GS matrix is created. This GS matrix is then transformed to an M x M correlation matrix by calculating the Pearson correlation between unassigned syllables. This correlation matrix is then transformed to a dissimilarity matrix by subtracting all correlations from one. The dissimilarity matrix is used as the input to the hierarchical clustering function, whose dendrogram output is, again, trimmed iteratively over a range of merging thresholds as described above.

Since no GS scores are calculated between session two syllables, the measure of cluster cohesiveness determined at each merging threshold is the proportion of the variance explained by the cluster eigensyllable, ranging from 0 to 1. As clusters are merged and greater variability is added to each cluster, the proportion of variance explained by each eigensyllable decreases. As with the initial clustering step, novel numbers of clusters that remain stable over multiple merging thresholds are returned along with the proportion of variance explained by each eigensyllable and the user must select the appropriate merging threshold. At this point, spectrograms are created and presented for validation. Upon arrival at an appropriate merging threshold, the second session syllable sequence is named by cluster assignment to generate a syntax string suitable for comparison with the first recording session.

Quantification of syntactical similarity

Syntactical similarity between two recordings (e.g. one recording session vs. another of the same animal) was determined by creating separate transition probability matrices for each song based on cluster assignments. The transition probability between two syllables is calculated by summing the total number of transitions between a leading syllable type and a following syllable

type, then dividing by the total number of transitions between the leading syllable type and all syllable types. This calculation is performed for all possible syllable-syllable pairs, including self-transitions.

Following construction of transition probability matrices, the transition behavior of the same syllable in both recording sessions is compared by calculating the Pearson correlation between corresponding rows. High correlation between rows indicates similar transition relationships for a given syllable in the two songs recordings being compared. To account for possible differences in frequency of syllable occurrence between the two recordings (e.g. the animal emits numerous renditions of vocalization 'A' in one recording session and very few in another), one minus the absolute difference in syllable type frequency (= the ratio of number of renditions of a given vocalization type to the total number of all vocalization types) is multiplied by the Pearson correlation for each corresponding row in the transition probability matrices. Finally, to obtain an overall score for syntactical similarity, the average Pearson correlation between corresponding rows of the transition probability tables, weighted by the difference in syllable type frequency, is calculated. It is worth noting, however, that correlation between individual rows can be informative to discover whether certain vocalization types' transition behavior is more responsible for driving an overall change in syntactical similarity.

In the event where one recording contains at least one vocalization type that is not present in the other (e.g. a novel vocalization emerges following an experimental manipulation), the transition probability matrices are modified so that the unique vocalization type is represented as a row and column of zeroes in the transition probability matrix for the session the novel syllable

type was not present in. Therefore, when correlations between rows in transition probability matrices are calculated, the unique syllable type results in a correlation of zero, providing a penalty when averaging all correlations. To quantify syllable transition similarity for only the syllable types that are present in both sessions, the averaged correlation only includes information for transitions between syllables found in both recording). In addition to the weighted scores, our method also returns the unweighted versions, which do not account for potential differences in syllable occurrence between recording sessions. The syntactical similarity metrics are summarized in Table 3-1.

$$Syntax\ Similarity = \sum_{i=1}^{n} \frac{cor(TPi_a, TPi_b) \bullet (1 - |fi_a - fi_b|)}{n}$$

n = number of syllable types; TP = transition probability; a, b = recording sessions; f = syllable frequency

Results

Overview: Semi-automated clustering of vocalizations

We present a method for the semi-automatic clustering of finch song syllables and mouse USVs through hierarchical clustering and automated dendrogram trimming. VEs in the form of zebra finch song syllables or mouse pup ultrasonic calls, were scored against themselves in a pairwise fashion to determine their acoustic similarity. The dimensionality of the resulting similarity matrix is limited only by the number of VEs that were recorded and used for input. This high degree of dimensionality provides greater specificity in grouping similar vocalizations, as compared to when clusters are based only on a finite number of acoustic features. The spectral co-similarity relationships between syllables are next subjected to hierarchical clustering, to generate a dendrogram, which is then trimmed into clusters using an automated tree-pruning algorithm. Originally developed for gene coexpression analyses, this tree-trimming algorithm has repeatedly yielded biologically meaningful clusters of genes from hierarchical trees[45]. Key advantages over other clustering methods include that the number of clusters (in this case, syllable or call types) is not dictated by the experimenter, providing for unbiased calculation of vocal repertoire. Following pruning of the dendrogram and determination of the number of syllable or call types, acoustic data for vocalizations of the same type is compiled and a syntax is generated. Vocalizations from subsequent recording sessions can then be compared to existing clusters, enabling both phonological and syntactical assessments across time, experimenters, laboratories, strains, genotypes or any other condition.

Validation of VoICE in birds

Zebra finch songs consist of multiple syllables that are repeated in a specific pattern to form motifs, the neuroethologically relevant unit of a song[120] (Figure 3-1A). To validate VoICE in birdsong analysis, we examined the first ~300 syllables sung on two separate days, seven days apart. 'Session A' comprised 308 syllables and 'Session B' comprised 310. Due to the stereotyped nature of adult song, we predicted that songs would retain their phonology and syntax over time; an outcome that would support the utility of VoICE. Syllables from the Session A were extracted using the "Explore and Score" module of Sound Analysis Pro[68] (SAP). Similarity scores between all syllables were calculated (Figure 3-2) and the resultant similarity matrix was imported and hierarchically clustered in R, resulting in the production of a dendrogram. The algorithm produced 54 unique clusters, which were merged to 8 final clusters by a guided procedure (Supplementary Text, Note 1), each representing a syllable in the motif (Figure 3-1B). For each cluster, an 'eigensyllable' was calculated to represent the syllable that best describes the variance within the cluster. The syllables in each cluster were correlated to the eigensyllable and ranked to determine overall homogeneity in the cluster. The syllable with the lowest correlation to the eigensyllable was visually inspected to ensure that all syllables were properly assigned to each cluster. The average correlation of the lowest ranked syllable to the eigensyllable across all clusters was 0.788, which captures the stereotypy of adult birdsong.

To test the expectation that phonology and syntax would be equivalent between the two sessions, syllables from Session B were assigned to clusters representing the Session A syllable types using a global similarity floor of 60 (Figure 3-1C)**.** Syllables from Session B were assigned

to Session A clusters using one of three possible outcomes: 1) 300 syllables were algorithmically determined to belong to a specific cluster ('assignment') resulting from a pairwise-comparisons post-hoc test following a statistically significant ANOVA result ($p<0.05$) for all clusters exceeding the global similarity floor; 2) 21 syllables were manually assigned to a cluster in the case of a non-significant ANOVA result; and 3) 0 syllables were deemed 'novel'. To test the validity of the novel syllable classification, 20 renditions of two syllables from a different songbird species, the Bengalese finch (*Lonchura striata domestica*), were subjected to the assignment procedure. As predicted, the 40 Bengalese finch syllables were deemed 'novel' and appropriately assigned to two new clusters (Figure 3-1C, 3-1D).

To compare sequential similarity between Sessions A and B, four syntax similarity scores were calculated (**Methods**), which can account for differences in syllable frequencies, and novel or omitted syllables. Comparison of Session A to Session B yielded syntactical similarity scores ~1.0, indicating a near perfect match between the two syntaxes (Figure 3-1E**,** 'unmodified'). In contrast, when 40 novel Bengalese finch syllables were randomly inserted into the Session B syntax (Figure 3-1E, 'modified'), the scores that penalize for the addition of novel syllable types dropped to ~0.75. Acoustic features were compared between clusters generated for Sessions A – B. Mean pitch and Weiner entropy, the latter being an acoustic measurement of syllable "noisiness," were similar ($p>0.05$, resampling independent mean differences) (Figure 3-1F).

As a second method of validation, we compared results from VoICE with those derived from hand counts and sorting through visual inspection by an experienced birdsong analyst (Supplementary Text, Note 2). Both analyses returned a similar number of syllable types (human:

n=8; VoICE: n=7). The human observer characterized a small number (3/1105; 0.6%) of syllables as a distinct syllable type that the computer (and a second human observer) did not find as categorically 'different enough' to result in a separate syllable type, which demonstrates the influence of human bias on sound categorization.

Quantification of deafening-induced song deterioration

Like humans, zebra finches require auditory feedback to maintain mature vocalizations, and the degradation of zebra finch song structure and syntax in the absence of hearing is well characterized[118,121]. To demonstrate the utility of VoICE in tracking changes to vocalizations, two adult zebra finches (>120d) were deafened. Song deterioration and syntax impairment were evaluated over a 4-month time frame. Representative spectrograms illustrate the stereotypy of mature zebra finch song (Figure 3-4A**,** pre-) and the variability in the time course of deafening-induced song changes (Figure 3-4A**,** post-). Initial clusters were assembled from a pre-deafening singing epoch (Figure 3-1A, top). Syllables from the first analyzed time point following deafening were then assigned to the pre-deafening clusters. For each subsequent time point, the first ~300 syllables from each day were assigned using the most recently clustered session (Figure 3-4B). As syllables degraded, the global similarity floor was manually lowered to 35 to enable continual assignment, reduce tiebreaking, and prevent novel syllable classification. After all time points were clustered, Wiener entropy (Figure 3-4C), and syntax similarity (Figure 3-4D) were examined (for additional acoustic measures, see Figure 3-5). As expected, syllable structure and syntax from a control bird (sham-deafened) were relatively unchanged throughout the recordings. In similarly-aged deafened birds, statistically significant changes to syllables were observed within 20 days

(one-way resampling ANOVA, multiple comparisons post-hoc Bonferroni corrected p-value < 0.05). In comparison, changes to the syllables of the sham-deafened bird were smaller, in a different direction, and occurred after ~80 days, possibly reflecting ongoing behavioral precision with aging. The songs of the two deafened birds deteriorated in different domains – one had significant decreases in the entropy of his syllables consistent with syllable degradation (Figure 3-4D, blue), whereas the other bird showed substantial decay in syntax (Figure 3-4D, red), but only minor phonological changes. Both phenomena have been previously observed following deafening in this species, supporting the ability of VoICE in capturing key facets of birdsong[118,121-124].

Validation in determining mouse ultrasonic vocal repertoires

To validate VoICE in the analysis of USVs, a 5-minute recording session from a C57BL/6J mouse pup (P7) was examined using manual classification of calls, the current standard, and by using VoICE. In rodents, isolation-induced USVs are retrieval calls emitted by pups when separated from their mother, representing an infant-mother vocal communicative behavior thought to be relevant to autism spectrum disorder (ASD)[125,126]. Recordings revealed narrowband vocalizations in the ultrasonic range (~40-120 kHz; Figure 3-6A). During manual scoring, a spectrogram of each call was generated in MATLAB, then assigned to one of 10 canonical call types[114] or to three miscellaneous categories. The same set of calls was then analyzed using VoICE. The dendrogram created after acoustic similarity scoring (Figure 3-3) and clustering was trimmed, then clusters that displayed an above-threshold level of homogeneity were described by calculation of an "eigencall". All calls within each of these clusters were then correlated to their respective cluster eigencall and the single call with the highest correlation was selected as a

representative of the entire cluster. Clusters were then classified by inspection of only these representatives. The clusters classified by a single representative call displayed a high level of within-cluster acoustic similarity (Figure 3-6B). Syllables from clusters with a sub-threshold level of homogeneity were classified individually.

VoICE almost exactly duplicated the call type distribution achieved by hand sorting, indicating that it can replicate the result achieved by the current analytical standard (Figure 3-6C, $p = 0.96$, resampling paired differences, Supplementary Text, Note 2). The small non-significant differences observed between manual and semi-automatic classification methods mainly resided in more downward and fewer harmonic call assignments.

Ultrasonic repertoire in Cntnap2 KO mice

The *Cntnap2* knockout mouse has been validated as one of the few mouse models of ASD with construct, face, and predictive validity[116], making it useful for study of ASD pathophysiology. *Cntnap2* knockout pups emit a reduced number of USVs when separated from the dam[116], however it is not known if this reduction is associated with abnormal spectral emission patterns. Other genetic mouse models of ASD have shown altered vocal repertoire[12], allowing us to hypothesize that VoICE will detect a similar finding in the *Cntnap2* knockout (KO) mouse. Therefore, to demonstrate proof of concept, we analyzed recordings of vocalizations from *Cntnap2* KO and wild-type (WT) littermates, obtained from heterozygous crossings. At postnatal day 7 (P7), calls were recorded for 5 minutes and then processed using VoICE. The reduced call number previously reported in KOs[116], was replicated here in a new cohort of animals (Figure 3-7A).

Beyond call numbers, using VoICE to conduct a comprehensive analysis of the vocal repertoire in these mouse pups allowed us to observe the hypothesized differences between the genotypes for multiple call types. Relative to WT, a smaller amount of the *Cntnap2* KO vocal repertoire was devoted to flat, frequency step, and harmonic calls and a greater amount to chevron, complex, downward, and triple calls ($p < 0.05$, See "Statistical representation of ultrasonic vocal repertoire difference" in Methods, Figure 3-7B, 3-7C). These differences replicated trends observed in a smaller pilot cohort of pups of the same genotypes recorded in 2008 (data not shown), speaking to the robustness of both the phenotype and the analytic method. By using VoICE, we were able to assess the similarity of vocal repertoire across animals by computing the Pearson correlation between raw call counts for each animal within the two genotypes. The within-KO correlation ($\rho = 0.60$) was greater than that of the within-WT ($\rho = 0.52$), suggesting a restricted repertoire across all KO mice (Figure 3-7D, 3-7E**,** $p = 0.0004$, resampling independent mean differences).

Since VoICE results in classification of all vocalizations to a canonical call type, we were able to apply the same syntax analysis metrics used in the study of bird songs to the mouse pup calls. To do so, we quantified the average weighted unpenalized syntactical similarity between all pups within each genotype, and found similar within-genotype syntactical relationships (KO = 0.42, WT = 0.39, Figure 3-7F), indicating that genotype has no effect on how similarly the animals order their calls. When compared across genotype, average syntactical similarity decreased (WT vs. KO = 0.33), an expected result given the difference in frequency of each call type between genotypes. Finally, we calculated syntax entropy scores, which describe the level of sequence variability with an animal's vocal repertoire. Low syntax entropy scores indicate that vocal

sequence is stereotyped (e.g. the occurrence of one call type very frequently precedes, or predicts, the occurrence of another call type). Syntax entropy scores did not differ between genotypes, indicating that genotype has no effect on how well the occurrence of one call type predicts the next ($p = 0.41$, resampling independent mean differences, Figure 3-7G).

Discussion

We here present a new methodology to empirically derive distinct vocal repertoires in avian and rodent species in a streamlined and unbiased manner (Figure 3-8). This is achieved by scoring acoustic similarity between individual song syllables or USVs and algorithmic trimming of a hierarchical cluster tree. We also provide a method for quantifying the syntactical similarity between bird songs and mouse USVs to assess the impact of experimental manipulation (e.g. auditory deprivation, gene knockout) on vocal behavior. These algorithms are not confined to zebra finch and mouse vocalizations, and can, in theory, be used to group VEs from any animal, so long as spectral co-similarity relationships can be appropriately quantified.

Our initial goal was to develop a method for grouping similar vocalizations together in the absence of input from the user, in order to generate a vocal syntax in an unbiased fashion. While an algorithm that intuitively trims dendrograms in an automated fashion should prove useful for accomplishing such a goal, user-defined thresholds that influence the number of clusters pruned from the tree still exist, prompting the need for empirical derivation of the "correct" number of syllable types in the animal's repertoire. In the analysis of finch songs, our novel clustering approach provides the investigator with the appropriate information to make an accurate and reproducible decision in deriving vocal repertoire extent without actually viewing the syllables as they are grouped, thus maintaining a high degree of impartiality. Upon completion of clustering, should the experimenter deem his or her chosen merging threshold to be incorrect (e.g. by viewing the cluster spectrograms and determining that clusters were inappropriately split or joined), the selection of a new merging threshold does not implicitly dictate the merging or splitting of a given

cluster, but instead allows a greater or lesser tolerance for variability within clusters (Supplementary Text, Note 1).

In contrast to the highly stereotyped vocalizations of adult finches that are shaped through development and actively maintained in adulthood in the presence of auditory feedback, mouse pup retrieval calls are considerably more variable on a rendition-to-rendition basis. This variability makes the subjective derivation of call "sameness" exceedingly difficult. While VoICE still operates within the canonical call categorization system, by calculating acoustic similarity relationships and clustering syllables before their classification, the possibility of determining call subtypes within the canonical call classifications becomes possible, reducing the amount of within-type variability when quantifying acoustic properties. For example, the calls in each of the algorithmically defined clusters displayed in Figure 3-3B all received a "downward" classification despite the obvious differences in frequency range, frequency contour, and duration across the clusters. Further, VoICE increases throughput by classifying entire clusters based on a single call mathematically determined to represent the greatest proportion of variance within the cluster. Accepting a greater level of variability when defining cluster cohesion, a determination that can be made based on experimental needs, can further increase the level of throughput.

Due to the algorithmic and mostly unsupervised grouping of VEs presented here, the quality of clustering is highly dependent on the quality of the input data and the measurement of spectral co-similarity scores. Should improvements be made in the scoring of similarity between vocalizations, the already high quality output of VoICE should increase. The extraction of songbird syllables and mouse calls is subject to user-defined segmentation parameters that both

positively and negatively influence the spectral scoring of each sound. The start/stop boundaries of individual song syllables are manually determined in order to provide discrete syllable .WAV files as input to SAP's "Similarity Batch." For USV analysis, calls are extracted using a user-defined amplitude threshold, which can result in slightly truncated or elongated sounds. Additionally, sounds separated by less than 10 msec are grouped as a single call (which may result in "double" or "triple" categorization), because it is unknown what silence duration reliably distinguishes two distinct calls.

In birdsong analysis, the application of VoICE is not limited to the assessment of deafening-induced song deterioration. Assignment of syllables from a pupil bird to its tutor could separately quantify learning in the phonological and syntactical domains. Other potential experiments include clustering song syllables during a control condition (e.g. before injection of a drug or virus) and then comparing subsequent recording sessions to control song clusters to quantify changes in syntax and phonology as a result of the treatment.

ASD, a pervasive developmental disorder, is multigenic in origin. One core endophenotype is a deficit in speech and language skills, which impairs social communication and well-being[127]. Characterizing the vocal behavior of animal models will be critical in evaluating the role of autism-susceptibility and other (e.g. FOXP2) genes implicated in speech and language learning in an ongoing effort to test therapeutics for social communication disorders. Implementing analysis techniques that have benefitted the birdsong field may prove invaluable in the assessment of USVs in genetically modified mice. Here we used VoICE to reveal an altered vocal repertoire in mice lacking the autism susceptibility gene, *Cntnap2*. Though songbirds are

excellent models for vocal learning and provide insight into the neural underpinnings of speech and language, it is currently more difficult to assess the underlying molecular determinants due to the challenge of generating and maintaining transgenic lines of birds[128]. Rodents are genetically tractable laboratory models, but the lack of standardization of USV analysis limits their efficacy in assessing genetic components related to vocalization. Our method provides an easy-to-use framework that can unify the analysis of both innate and learned vocal signals to provide insight into the genetic and physiological mechanisms that comprise vocal communication.

Supplementary Text

Introduction to Supplementary Notes

The ideal system for clustering songbird syllables is an errorless and unbiased observer, which does not exist. Further, hand parsing of data is not feasible when considering thousands of syllables. Nevertheless, to evaluate VoICE, an hour's worth of song recordings from a zebra finch were manually clustered by an experimenter familiar with song analysis but blinded to the analytical goal, then the same set of recordings were passed through the VoICE pipeline by another experimenter familiar with the procedure.

The human observer found 1122 syllables in the hour of songs. The experimenter using VoICE found 1105 syllables in the same hour, with the minor discrepancy likely due to difference of opinion between experimenters as to the initiation and termination of song bout boundaries. Still, both experimenters largely considered the same song content in their respective analyses.

Supplementary Note 1: Selecting a Merging Threshold

After construction of the M x M distance matrix as outlined in Online Methods, a dendrogram was created and trimmed, resulting in the creation of 65 unique clusters (Figure 3-9A, 'unmerged'). The tree trimming procedure was then iteratively repeated and the merging threshold decreased from 1 to 0 by steps of 0.01 with each iteration. Upon completion of iterative tree trimming, the cluster numbers that remained stable over at least two merging thresholds were set aside for further analysis (Figure 3-9B).

At each stable merging threshold, the user is then presented with the IGS for each cluster and the number of syllables present in that cluster. Ultimately, the user must determine the correct merging threshold by weighing the balance between the number and size of clusters, IGS, the number of merging thresholds over which the cluster n remained constant. Six unique cluster definitions were stable over multiple merging thresholds, narrowing the possible number of syllable types to a range between five and 10 (Table 3-2).

The merging thresholds close to 1 resulted in high cluster n, coincident with the existence of fewer, smaller clusters. The presence of multiple distinct syllable types with very few renditions each in an adult zebra finch's song is unlikely, suggesting that utilizing a very high merging threshold is too strict and results in under-merging of clusters. Conversely, merging thresholds close to 0 resulted in a low cluster n with relatively low IGS due to increased heterogeneity within the cluster. Based on these observations, merging thresholds at the extremes of the spectrum were removed from consideration (0.94, 0.92, 0.58, 0.39, 0.35).

Finally, merging thresholds of 0.9 (n=8 clusters) and 0.79 (n=7 clusters) were considered. When the threshold is lowered to 0.79, the cyan cluster (n=3 syllables, IGS=85.69) is merged into the purple (n=227 syllables, IGS=82.12) cluster. Following this merge, the purple intracluster identity decreases to 80.43 (Table 3-2), indicative of an average score of 93 for similarity, accuracy, and sequential match between all syllables in the cluster. When the hierarchical tree was trimmed using the 0.79 merging threshold, seven clusters were generated (Figure 3-9A, 'merged'). Manual error checking of clusters revealed that two syllables were placed in the incorrect cluster, resulting in an error rate of ~0.18%.

Supplementary Note 2: Comparison to Human Scoring

A discrepancy occurred in the number of syllable types present as determined by the experimenter manually clustering the syllables (n = 8, Figure 3-10A) versus using the threshold determined by iterative trimming of the hierarchical tree (n = 7, Figure 3-10B). When sorted by hand, the syllable type determined by clustering, purple, was subdivided into syllables B (n=224 syllables) and H (n=7 syllables) (Figure 3-10). The merge in question eliminates the presence of a cluster containing only three syllables in a total of 1105. It is possible that an adult zebra finch could sing a distinct syllable type as ~0.6% of its song, but the more parsimonious interpretation is that syllable H, while somewhat dissimilar from syllable B, is still of the same "type." Indeed, syllables B and H are largely similar: both are of approximately the same pitch (median = 375.5 hz and 402 hz, respectively) and duration (median = 55.87 msec for both), though syllable B is slightly more frequency modulated (median = 40.5 vs. 22.15). Therefore, for the purpose of comparing syntax scores between the manual vs. semi-automated approach, all syllables scored as "H" were renamed to "B." When considering the syllables for the purposes of acoustic analyses, however, one can opt to deem syllable H as a subtype of B (e.g. B_i) and consider their acoustic properties separately.

To compare the two methods for quantifying song syntax, transition probability tables were created and these methods resulted in very similar scores, with the advantages of VoICE being faster in the processing of larger data sets and introducing less experimenter bias. There were marginal differences found between the two methods and transitions that were present in one analytical method that did not exist in the other were inspected more closely (Table 3-3).

Only one transition absent in the hand sorting of syllables was present when syllables were clustered using VoICE. This amounted to a single yellow to purple transition. This discrepancy was potentially attributable to one of two possibilities: an error resulting from the procedure or the clustering analysis including a syllable that was not deemed part of a song bout by the experimenter sorting syllables by hand. The latter proved to be true as the number of syllables from the specific song-recording file found to contain the yellow-purple transition by the clustering procedure was 13 while manual scoring included only 11, illustrating and accounting for the transition probability discrepancy between the two analyses (Figure 3-11).

The weighted unpenalized syntactical similarity between transition probability matrices created from the semi-automated clustering results and the data scored by hand was 0.9994, indicating nearly identical syntaxes were identified by the two scoring methods.

Tables

Table 3-1: Summary of syntax similarity scores and description of transition behavior quantified by each metric

Unweighted	Weighted	Unpenalized	Penalized	Description
x		x		Ignores vocalization frequency and novel vocalizations between sessions
x			x	Ignores vocalization frequency, accounts for novel vocalizations between sessions
	x	x		Accounts for vocalization frequency, ignores novel vocalizations between sessions
	x		x	Accounts for vocalization frequency and novel vocalizations between sessions

Table 3-2: Number of clusters, number of syllables in each cluster (n_{syl}) and IGS at the first merging threshold that resulted in a stable cluster n at over at least two merging threshold changes

Threshold	Cluster ID	n_{syl}	IGS
0.94	red	128	83.6
(n = 10)	orange	221	81.8
	green	221	86.0
	blue	221	82.6
	yellow	38	85.1
	pink	45	70.4
	purple	227	82.1
	cyan	3	85.7
	magenta	2	53.8
	tapioca	2	35.5
0.92	red	128	83.6
(n = 9)	orange	221	81.8
	green	221	86.0
	blue	221	82.6
	yellow	38	85.1

	pink	47	65.7
	purple	227	82.1
	cyan	3	85.7
	magenta	2	53.8
0.9	red	128	83.6
(n = 8)	orange	221	81.8
	green	221	86.0
	blue	221	82.6
	yellow	38	85.1
	pink	49	63.2
	purple	227	82.1
	cyan	3	85.7
0.79*	red	128	83.6
(n = 7)	orange	221	81.8
	green	221	86.0
	blue	221	82.6
	yellow	38	85.1

	pink	49	63.2
	purple	230	80.4
0.58	red	128	83.6
(n = 6)	orange	221	81.8
	green	221	86.0
	blue	221	82.6
	yellow	38	85.1
	pink	279	64.3
0.39	red	407	52.0
(n = 5)	orange	221	81.8
	green	221	86.0
	blue	221	82.6
	yellow	38	85.1
0.32	red	407	52.0
(n = 4)	orange	221	81.8
	green	259	69.6

blue	221	82.6

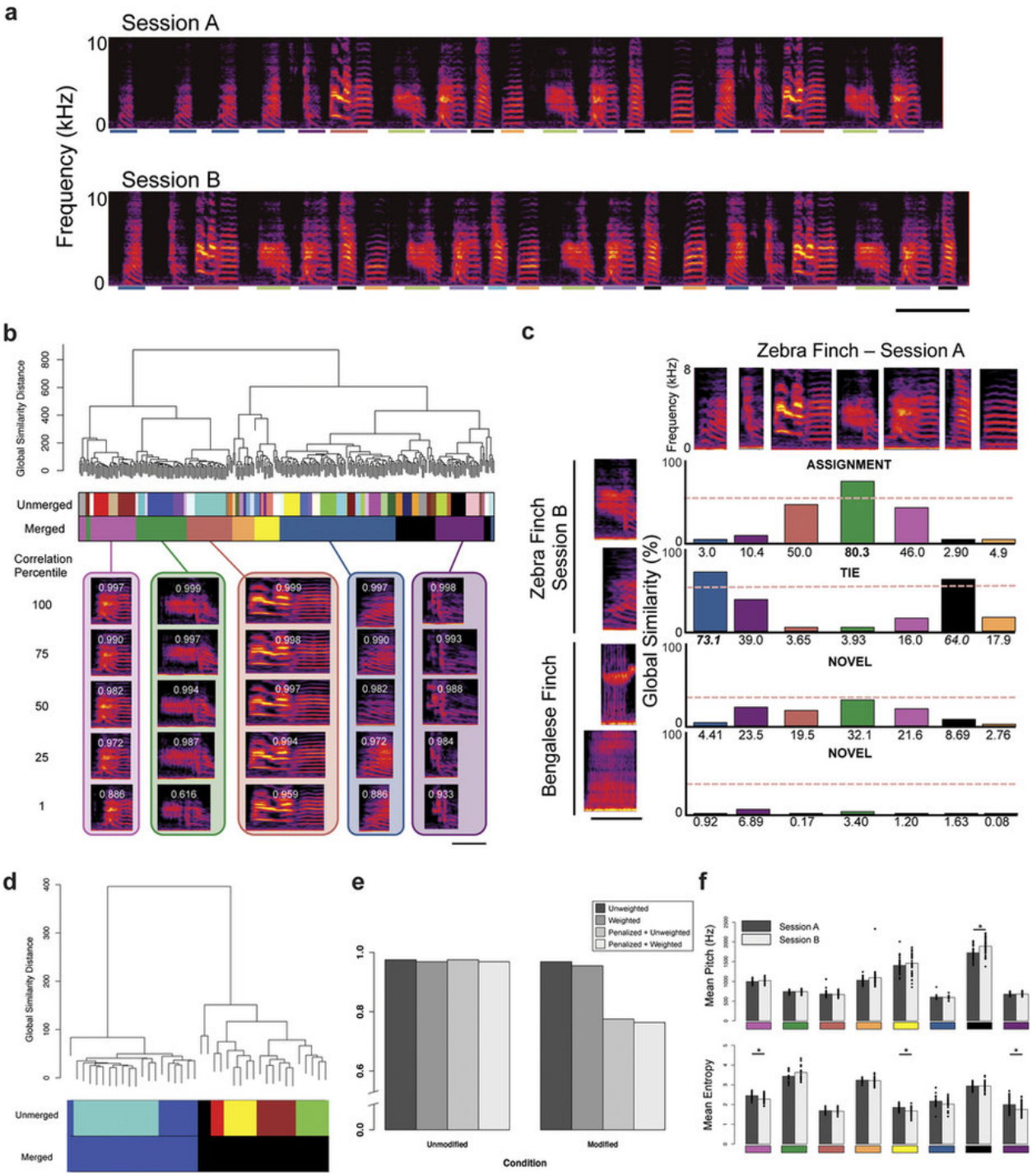
Asterisk denotes the merging threshold chosen and illustrated in Figure 3-9 ('merged').

Table 3-3: Comparison of transition probabilities between VoICE (top) and manual scoring by an investigator (bottom)

Method	Lead Syllable	Following Syllable						
		Red	Purple	Green	Blue	Orange	Pink	Yellow
VoICE	Red	5.5	94.5					
	Purple	3.9		96.1				
	Green				100.0			
	Blue					100.0		
	Orange	11.4	49.1				22.3	17.3
	Pink	100.0						
	Yellow	97.4	2.6*					
Manual	Red	5.4	94.6					
	Purple	3.0		97.0				
	Green				100.0			
	Blue					100.0		
	Orange	11.7	48.4				22.4	17.5
	Pink	100.0						
	Yellow	100.0						

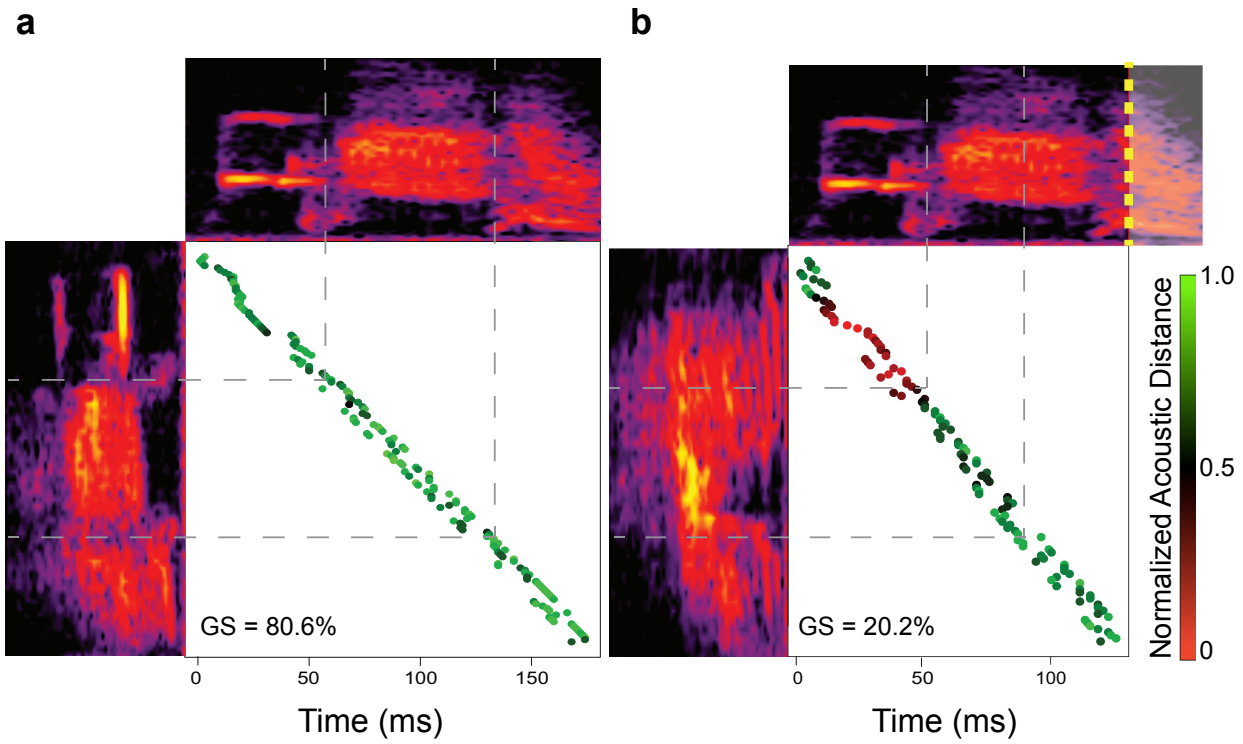
Figures

Figure 3-1: Assignment and quantification of clustered birdsong syllables



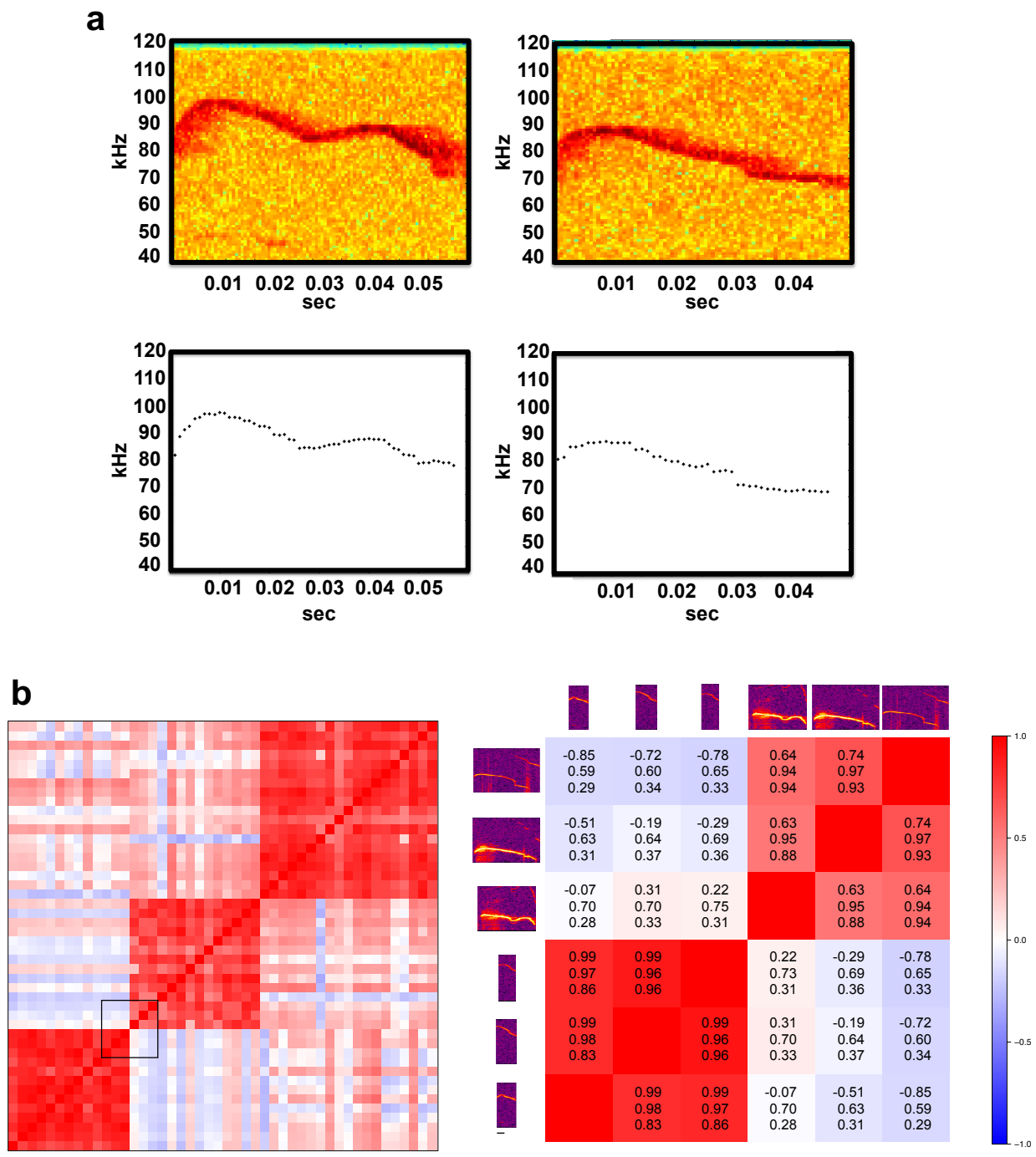
a) Mature zebra finches (>120d) sing stereotyped song composed of repeated syllables that form motifs that form bouts. Shown are two song bouts sung by the same adult bird during two recording epochs ('Session A' and 'Session B'). (Scale bar = 250 msec.) **b)** Dendrogram plots global similarity distance between leaves (syllables) and was generated following spectral similarity scoring. Beneath the branches, clusters before (Unmerged) and after merging (Merged) are denoted by color bands. Representative syllables from merged clusters are illustrated at descending percentiles following correlation of each cluster member to the cluster eigensyllable. The Pearson's rho for the correlation between each syllable and its eigensyllable are displayed in white. **c)** During assignment, one of three possible outcomes for each syllable occurs: automatic assignment to a cluster (ASSIGNMENT), manual assignment in a tiebreaking procedure when statistically similar to two clusters (TIE), or categorization as novel (NOVEL). Artificially introduced syllables from a Bengalese finch did not pass a global similarity floor and are accurately deemed 'novel'. Bars indicate the mean percentage global similarity between the syllable and each cluster. **d)** The two artificially introduced syllables from a Bengalese finch, are, upon merging (Merged), appropriately assigned to two novel clusters. **e)** Syntaxes are highly similar between recording sessions, regardless of metric used for scoring (left, 'unmodified') but the artificial introduction of novel syllables to the second recording session reduces similarity when using a metric that penalizes for novel syllables (right, 'modified'). **f)** Pitch (top) and entropy (bottom) are largely unchanged between recording sessions. (* = $p<0.05$, resampling independent mean differences. Cluster colors are consistent throughout. Scale bars = 50 msec.)

Figure 3-2: Zebra finch acoustic similarity scoring



Similarity scores are determined by averaging the millisecond-by-millisecond Euclidean distance of four acoustic features: pitch, Wiener entropy, frequency modulation, and goodness of pitch. **a)** Visually alike syllables are highly similar at each millisecond (green dots), and **(b)** distinct syllables are more dissimilar (red dots). The global similarity score (GS), which is partially determined by differences in syllable duration, for each pair of syllables is displayed in the lower left-hand corner of each plot. (Spectrogram frequency axis range 0 to 10 kHz.

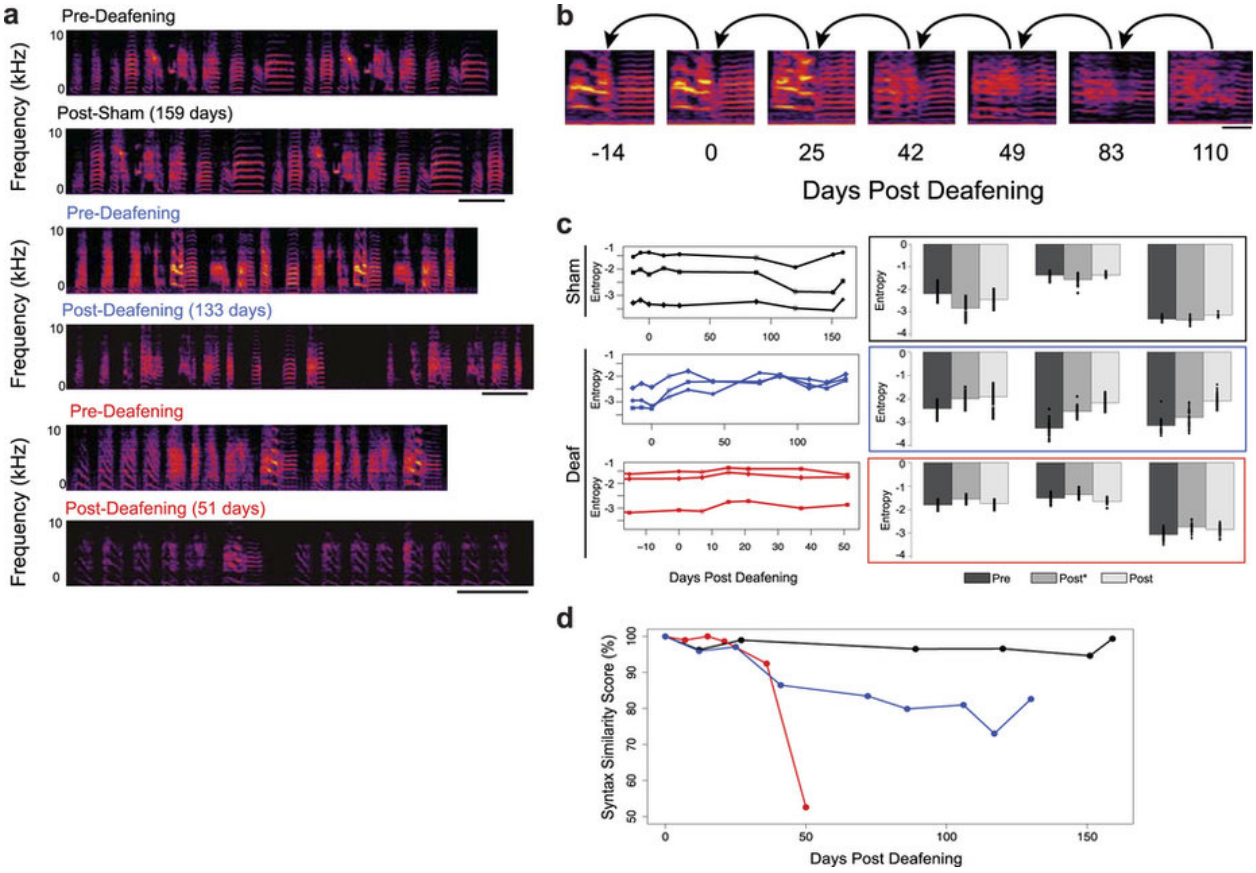
Figure 3-3: USV similarity scoring



a) Individual USV spectrograms (top) are transformed to frequency contours summarizing 0.9 ms windows before similarity scoring (bottom). **b)** Exemplar USV weighted correlation matrix used as input to hierarchical clustering algorithm represented as a heatmap (left) and inset, black square

(right), illustrating actual syllables and their pitch correlation, pitch difference, and temporal overlap scores (top to bottom, respectively). Rows and columns in the heatmap represent calls from one animal's recording session. The indices represent the spectral similarity scores between each pair of calls. Three clusters automatically defined by the tree-trimming algorithm were used as exemplars. (Scale bar = 10 msec.)

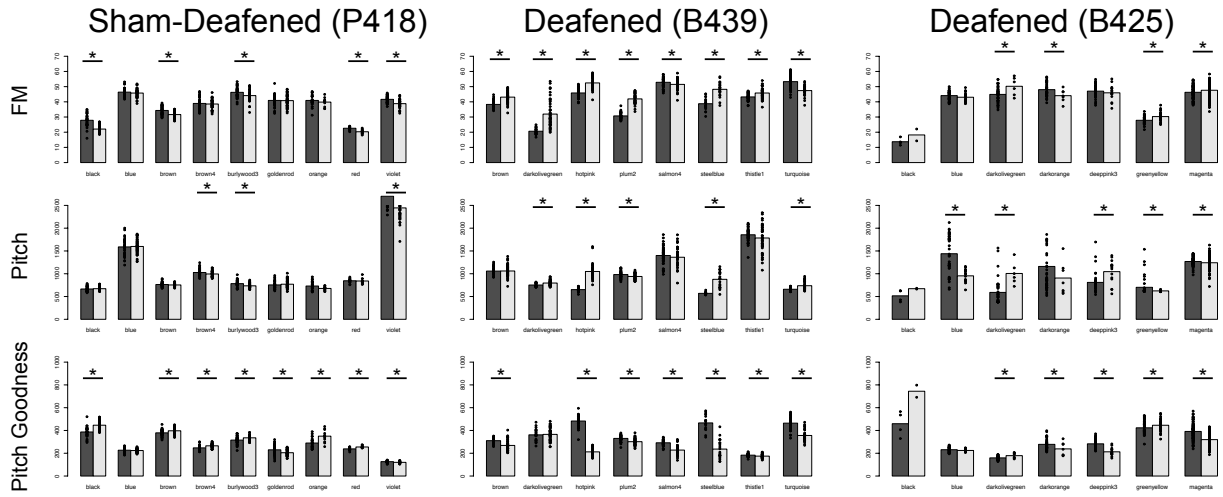
Figure 3-4: VoICE detects deafening-induced alterations in song phonology and syntax



a) Spectrograms reveal song deterioration in deafened, but not sham-deafened, birds. **b)** Syllables are assigned in a temporally-reversed serial manner to account for ongoing changes in syllable structure. **c)** Syllable entropy, a measure of spectral 'noise', increases in a majority of syllables after deafening. Asterisks denote statistically significant changes from before surgery (left). Bar plots represent Pre (Day 0) vs. Post* (the first day statistically significantly different from 'Pre') vs. Post (the last analyzed day) recordings. Each symbol and line (left) and its corresponding pair of bars (right) represent a syllable cluster (right). (One-way resampling ANOVA, multiple comparisons post-hoc Bonferroni corrected p-value < 0.05) **d)** Syntax similarity to pre-surgery

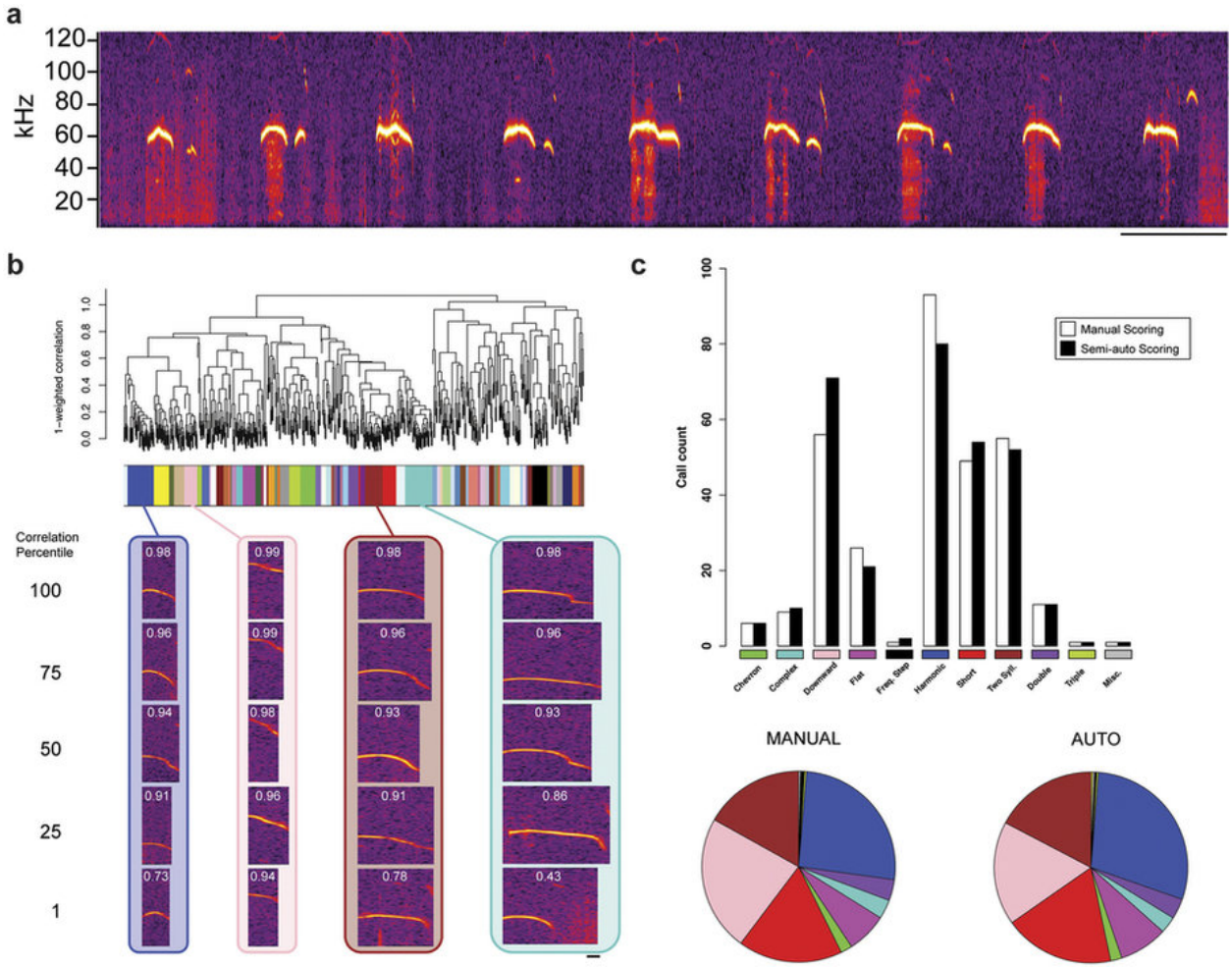
decreases following deafening. (Black = sham; blue, red = deaf, * = $p<0.05$ resampling independent mean differences. Scale bars = 250 msec in (a) and (b).)

Figure 3-5: Quantification of multiple acoustic features before and following deafening



The mean **(a)** frequency modulation (FM), **(b)** Pitch, and **(c)** Pitch Goodness for each cluster of a sham-deafened bird and two deafened birds reveals the consequences of auditory manipulation in each bird. Each dot represents a single syllable. (* = $p<0.05$, resampling independent mean differences.)

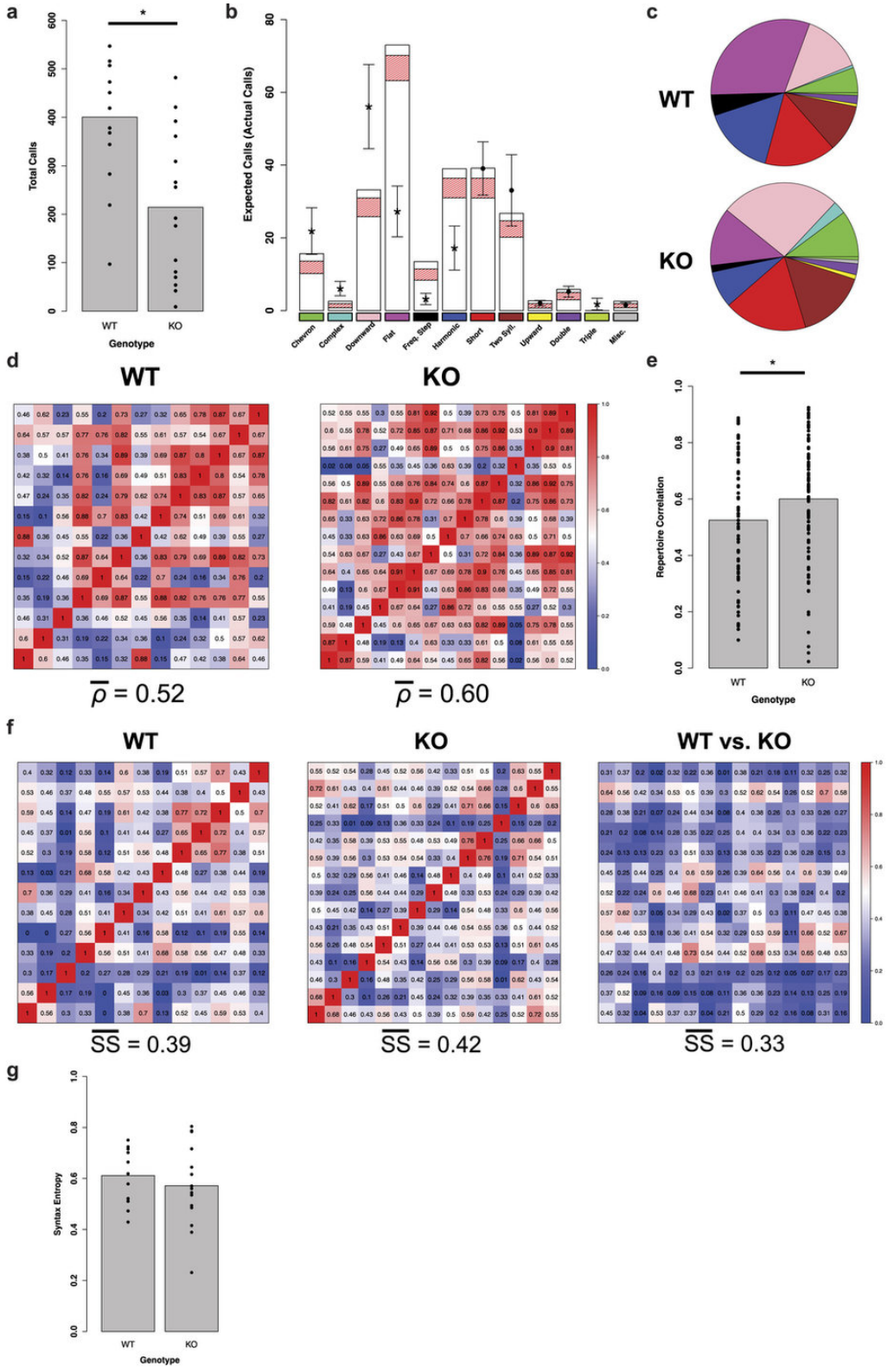
Figure 3-6: Validation of USV technique and comparison to manual classification standard



a) Exemplar USVs from a mouse on the C57BL/6J background at P7. (Scale bar = 200 msec.) **b)** A dendrogram generated following spectral similarity scoring of USVs where calls are represented as leaves and branch points indicate the difference in weighted correlation between leaves. Beneath the branches, clusters automatically determined by the tree-trimming algorithm are denoted by unique color bands and illustrated by representatives at descending percentiles following correlation of each cluster member to the cluster eigencall. The Pearson's rho for the correlation between each syllable and the eigencall are displayed in white. **c)** Bar plots indicate the count of each call type when the classification is performed manually (white) or using VoICE (black). Pie

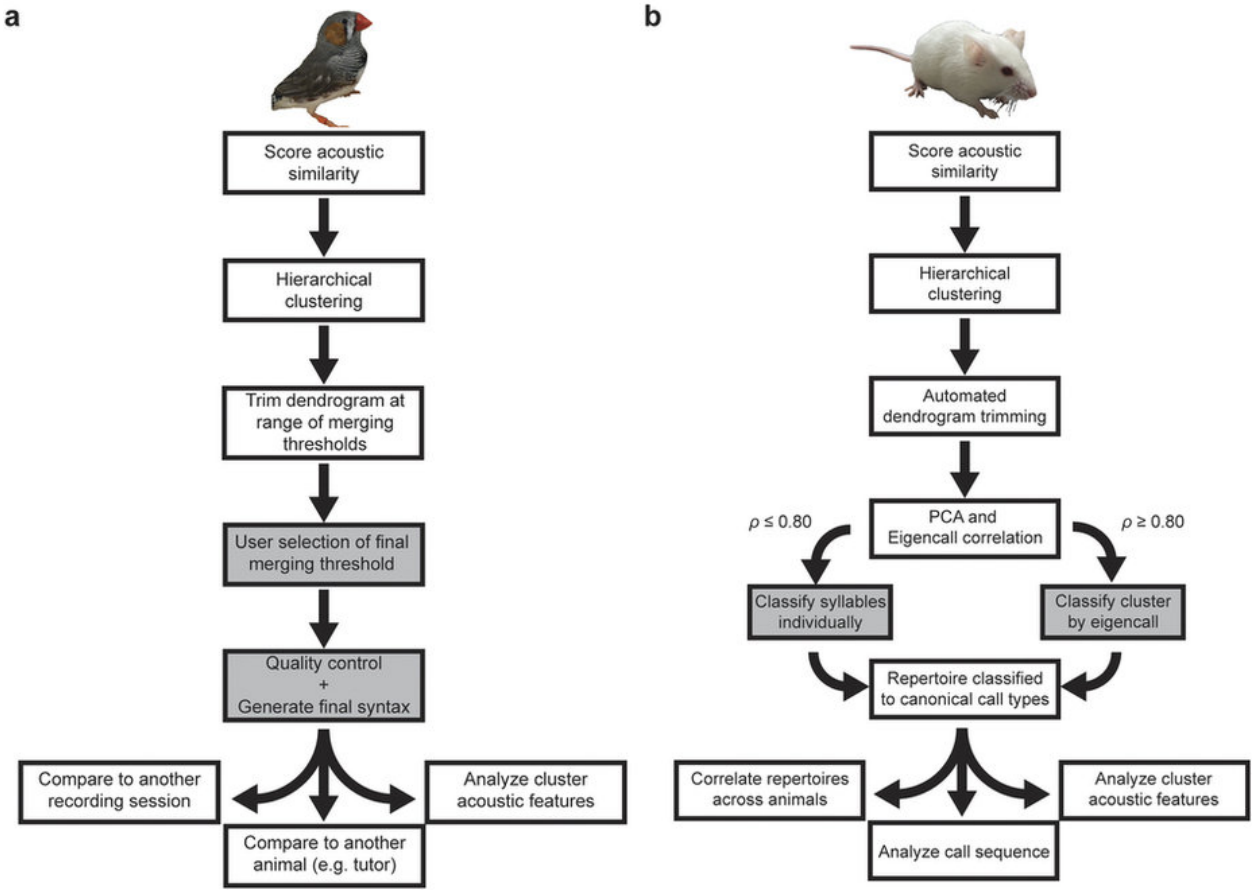
charts, right, illustrate the percentage distribution of each call type for the same animal's repertoire as determined by manual sorting or using VoICE (Scale bar = 10 msec.)

Figure 3-7: Deletion of *Cntnap2* results in altered vocal phenotype



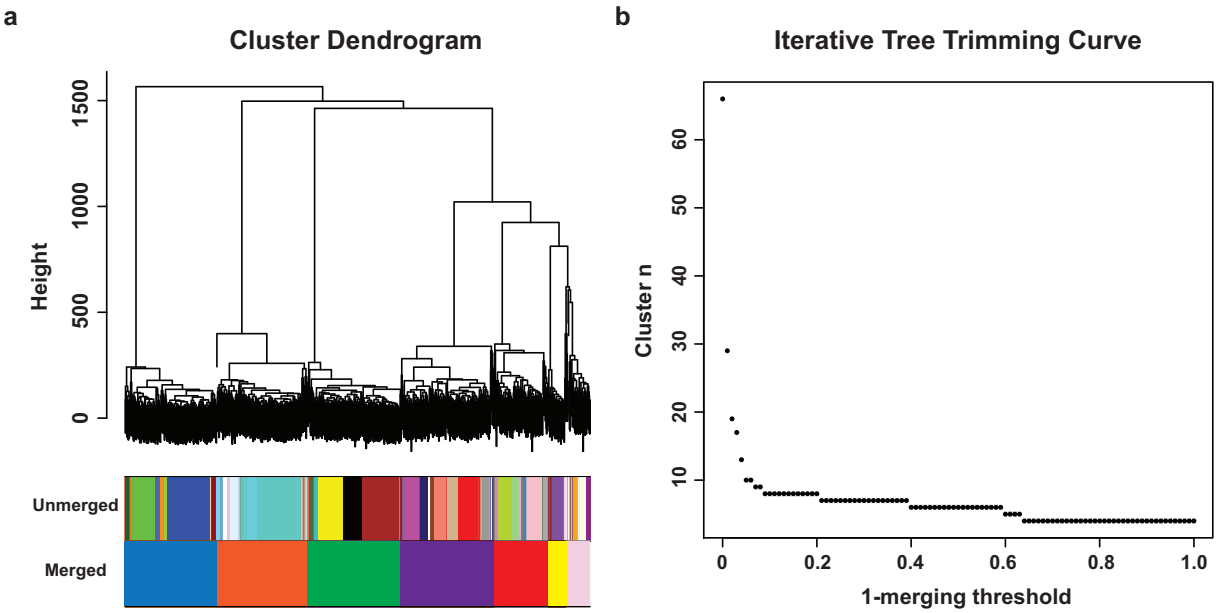
a) Mouse pups lacking *Cntnap2* (n=15) do not call as much as WT littermates (n=13) (* = $p<0.05$, resampling independent mean differences) **b)** Expected counts of each call type (bars) generated from resampled WT data and 95% confidence intervals (red cross-hatch) reveal significant differences when compared to actual KO call counts, represented by overlaid points. Average counts of actual KO calls are represented as asterisks where $p<0.05$. Error bars denote ± s.e.m. **c)** Pie charts display the distribution of each call type in WT and KO animals. (Color scheme denoted beneath bars in (b) **d)** Heatmaps denote the correlation of repertoire within each genotype. KO animals show an intragenotype correlation greater than that of WT. Rows and columns represent animals, and indices are repertoire correlations between them. **e)** Repertoire correlation is significantly greater within the KO genotype. **f**) Heatmaps of the within- and across-genotype weighted unpenalized syntactical similarity scores show no within-genotype difference in syntax similarity. Rows and columns represent animals, and indices are syntax similarity scores between them. **g**) Syntax entropy scores (a measure of call transition variability) within each genotype are similar.

Figure 3-8: Summary of procedures



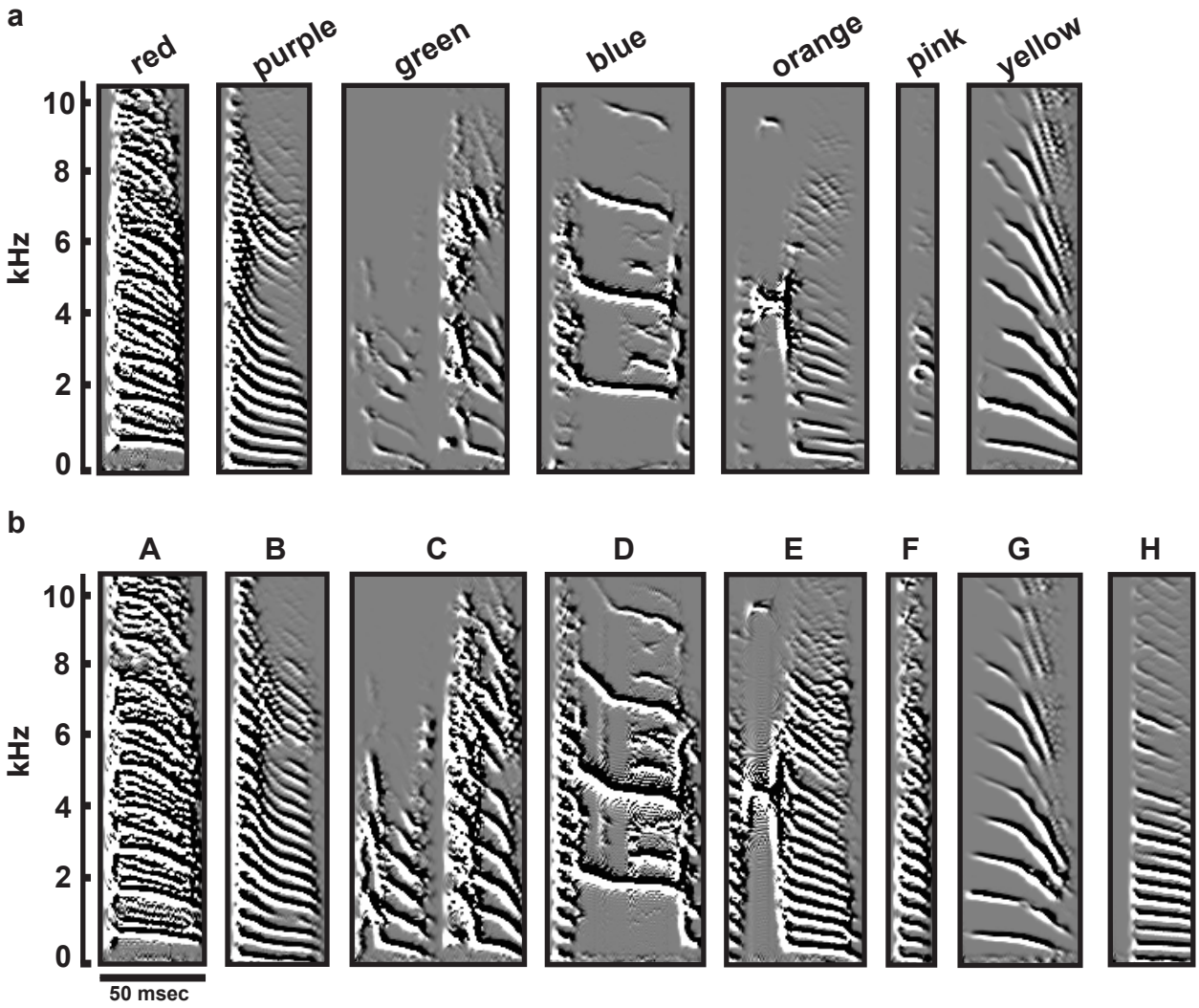
Flow charts describe the analytical pipeline for **(a)** zebra finch and **(b)** mouse USV analyses. Steps at which user input occurs are shaded in gray.

Figure 3-9: Detailed clustering results



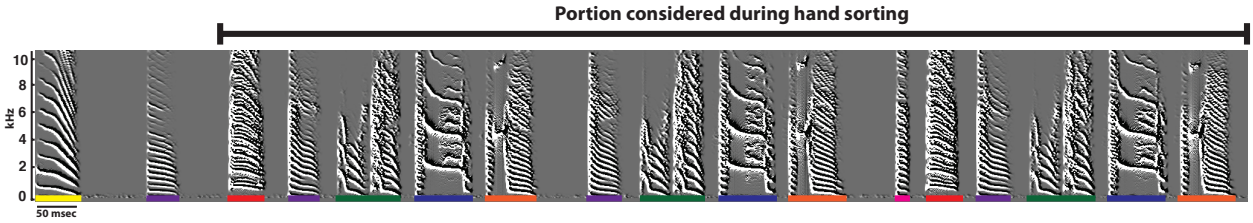
a) A dendrogram generated for 1105 syllables recorded during one hour of singing. 'Unmerged' colors represent the most divisive trim by the automated tree-trimming algorithm. 'Merged' colors are groups following guided dendrogram trimming, consisting of cluster ns that remained stable over multiple merging thresholds in **(b)**.

Figure 3-10: Comparison of unique clusters determined by different methods



a) Seven syllable clusters were determined by guided dendrogram trimming using VoICE. **b)** Eight clusters were assessed by the experimenter scoring by hand.

Figure 3-11: Determination of transition discrepancy between VoICE and human scoring results from difference of opinion between the onset of a singing bout



Colors indicate cluster assignments as determined by VoICE.

Acknowledgements

We thank AT Hilliard (Stanford) and S Horvath (University of California, Los Angeles (UCLA)) for consultation on the application of clustering methods, A Garfinkel (UCLA) for statistical advice, and ER Fraley (UCLA) for feedback about mouse vocalizations. We also thank BS Abrahams (Albert Einstein College of Medicine) for collecting pilot data.

Epilogue

The launch of VoICE provided a convenient tool for researchers to quickly and reliably quantify the vocalizations of birds and rodents. It also fostered opportunities for collaboration with other research groups eager to characterize the vocal consequence of genomic manipulations in their model species of interest. These collaborations lead to publication of multiple publications. Those most focused on vocal behavior are presented as complete manuscripts in Appendices 1-3. An additional publication wherein the phenotype resulting from knockout of the autism susceptibility gene, *JAKMIP1*, upon which I am an author, was an early application of VoICE. Since this publication contains substantial characterization of the non-vocal phenotype consequent of *JAKMIP1* knockout that I did not participate in and is not relevant to my research goals as a graduate student, I present here the portion of this manuscript relevant to my efforts as an epilogue to the first chapter.

JAKMIP1, a Novel Regulator of Neuronal Translation, Modulates Synaptic Function and Autistic-like Behaviors in Mouse

Jamee M. Berg, Changhoon Lee, Leslie Chen, Laurie Galvan, Carlos Cepeda, Jane Y. Chen, Olga Peñagarikano, Jason L. Stein, Alvin Li, Asami Oguro-Ando, Jeremy A. Miller, Ajay A. Vashisht, Mary E. Starks, Elyse P. Kite, Eric Tam, Amos Gdalyahu, Noor B. Al-Sharif, Zachary D. Burkett, Stephanie A. White, Scott C. Fears, Michael S. Levine, James A. Wohlschlegel, Daniel H. Geschwind

JAKMIP1 KO Mice Display ASD-Associated Behaviors

Recent work, including knockout of the eukaryotic translation initiation factor 4E-binding protein 2 (4E-BP2) [129] and overexpression of the eukaryotic translation initiation factor 4E (eIF4E) [130], suggests that disrupting mRNA translation can lead to ASD-like behaviors in the mouse. We developed a line of knockout mice and characterized multiple ASD-related behavioral domains including repetitive behavior, social abnormalities, and altered vocal communication, as well as other associated phenotypes including increased impulsivity and motor abnormalities in these *Jakmip1* KO mice.

Jakmip1 KO mice showed striking motor stereotypies during home-cage behavior including significantly increased grooming and repetitive jumping behavior with over 90% penetrance (Figure 3-12A and 3-12B). *Jakmip1* KO mice also show increased perseveration in the T maze, displaying significantly fewer alterations than WT mice (Figure 3-12C). These behaviors are considered mouse analogs of restrictive and repetitive behaviors observed in ASD [126]. Probing social behavior with the three-chamber task also revealed social dysfunction in *Jakmip1* KO mice (Figure 3-12D and 3-12E). This social deficit was not due to disrupted olfaction, as *Jakmip1* KO mice showed normal olfaction in the buried food test.

To determine if *Jakmip1* KO mice display disruptions in vocal communication, we tested ultrasonic vocalizations in postnatal WT and KO mice upon separation from their mother. These distress calls are thought to model early vocalization abnormalities in ASD infants and are observed in several mouse models of ASD, including *Fmr1* KO mice [114,116,131,132]. We

found that *Jakmip1* KO mice showed a significant increase in vocalizations and various call types across all time points (Figure 3-12F) not related to weight.

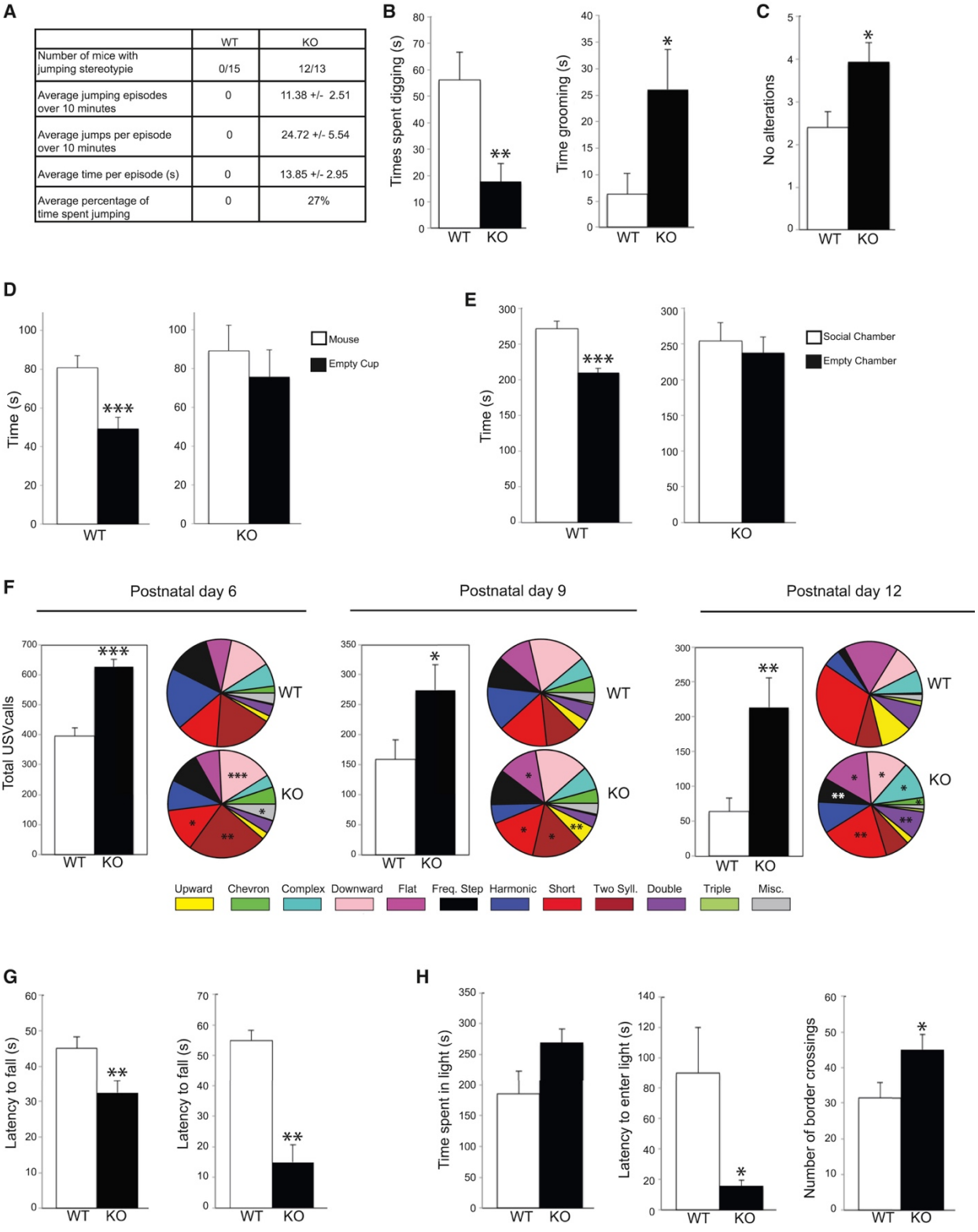
Jakmip1 KO mice also displayed significant deficits in motor function and impulsivity, which, although not core diagnostic features of ASD, are frequently observed in patients, including those with fragile X syndrome [133,134]. *Jakmip1* KO mice performed poorly on both the accelerating rotarod and wire hang test, indicating abnormal motor function (Figure 3-11G). In the light-dark box test for anxiety, *Jakmip1* KO mice showed a trend toward spending more time in the light compartment compared to the dark compartment, a significant reduction of latency to enter the light compartment, as well as increased border crossings (Figure 3-11H), suggestive of either reduced anxiety or increased impulsivity [135].

To test for learning impairments, we conducted auditory fear conditioning, which measures hippocampal and amygdala-dependent learning [136,137]. *Jakmip1* KO mice were able to learn to freeze to a tone after tone-shock pairings during the acquisition phase of fear conditioning but showed significant decreases in freezing during the second and third intertone intervals as compared to WT mice not due to impaired sensitivity to shock stimuli. *Jakmip1* KO mice showed normal context dependent fear conditioning, indicative of preserved hippocampal-dependent learning. They also performed normally in a test of generalized fear assessment when placed in a new. However, *Jakmip1* KO mice displayed decreased noise-cued fear response, suggesting disrupted amygdala/auditory pathway function[136].

Marble burying is a naturalistic repetitive behavior that is often reduced in WT C57BL/6 mice treated with anxiolytics or selective serotonin reuptake inhibitors [138,139]. *Jakmip1* KO

mice buried fewer marbles than WT mice, consistent with the increased impulsivity/decreased anxiety suggested by their performance in the light-dark box. They also showed decreased digging, another form of naturalistic repetitive behavior (Figure 3-11B). *Jakmip1* KO mice performed indistinguishably from WT mice on tests of sensory acuity and nesting and showed normal open-field activity.

Figure 3-12: *Jakmip1* Loss Leads to ASD-Associated Behaviors



	WT	KO
Number of mice with jumping stereotypie	0/15	12/13
Average jumping episodes over 10 minutes	0	11.38 +/- 2.51
Average jumps per episode over 10 minutes	0	24.72 +/- 5.54
Average time per episode (s)	0	13.85 +/- 2.95
Average percentage of time spent jumping	0	27%

a and b) *Jakmip1* KO mice show repetitive and perseverative behavior in the home-cage behavior test. **a)** Characteristics of repetitive jumping stereotypy in *Jakmip1* KO mice. Rightmost column displays mean ± SEM. **b)** Time spent digging (left) or grooming (right) within a 10-min period. p values were calculated using a two-tailed unpaired t test (time spent digging, p = 0.0063; time spent grooming, p = 0.025; WT, n = 15; KO, n = 13). **c)** T-maze spontaneous alternation test. Number of alterations are shown. p value was calculated using a two-tailed unpaired t test (WT, n = 15; KO, n = 13; p = 0.015). **d and e)** *Jakmip1* KO mice show impaired social behavior in the three-chamber social test. **d)** Time spent sniffing a sex-matched novel mouse or an empty cup over a 10-min period. p values were calculated using a two-tailed paired t test (WT, n = 15; KO, n = 13; WT, p = 0.00051; KO, p = 0.51). **e)** Time spent in the social chamber containing a novel mouse or in the chamber containing an empty cup. p values were calculated using a two-tailed paired t test (WT, n = 15; KO, n = 13: WT, p = 0.00064; KO, p = 0.73). **f)** Number of USV calls emitted from postnatal day 6, 9, and 12 pups after being separated from their mother (5-min period). Pie charts to the right of graphs show call type distribution. Statistical significance is denoted on the KO pie charts with all significant call types increased in KO mice (P6 [WT, n = 15; KO, n = 15; p = 1.2E-6], P9 [WT, n = 17; KO, n = 17; p = 0.038, and P12 [WT, n = 15; KO, n = 15; p = 0.0034]). p values were calculated using a two-tailed unpaired t test. **g)** *Jakmip1* KO mice show impaired motor coordination and decreased strength. Left: accelerating rotarod. y axis is latency to fall from the rotarod. Maximum time of trial is 180 s. p values were calculated using a two-tailed unpaired t test (WT, n = 15; KO, n = 13; p = 0.011). Right: wire-hang test. y axis is latency to fall from an inverted wire cage lid. Maximum time of trial is 60 s. p values were calculated using the non-parametric Mann-Whitney test (WT, n = 9; KO, n = 7; p = 0.002). **h)** *Jakmip1* KO

mice show increased impulsivity/reduced anxiety in the light-dark box test. p values were calculated using a two-tailed unpaired t test. Left: time spent in the bright compartment over a 10-min period (WT, n = 15; KO, n = 13; p = 0.075). Middle: time before the mouse first enters the bright compartment (WT, n = 15; KO, n = 13; p = 0.029). Right: number of times the mouse crosses compartments over a 10-min period (WT, n = 15; KO, n = 13; p = 0.034). Data are shown as mean ± SEM.

Chapter 4: Conclusion & Future Directions

The broad goal of my research has been to identify the molecular basis for the complex behavior of learned vocalization, a key subcomponent of spoken language, in the zebra finch model system. In Chapter 1 I describe the songbird as a model system for studying the molecular basis for vocalization and communication, then highlight the key historical findings and more modern approaches in this ethologically relevant model system that have provided insight to the mechanisms underlying this behavior. Of particular note, a study in the White lab which concluded upon my arrival was an application of network theory to the transcriptome in order to relate gene coexpression patterns instead of absolute expression levels to the act of producing a learned vocalization. This study provided an important basis for my research and served as a frame of reference for my own results collected from the juvenile/actively learning songbird.

I present the main findings of my graduate study in the second chapter, where transcriptional networks positively correlated to vocal learning were discovered and described. The most exciting finding is that these learning-related coexpression patterns are very poorly preserved in Area X of adult/no longer learning zebra finches, suggesting that aging out of the critical period in which songs are changing may drive an alteration of the transcriptome that prevents vocal learning from occurring. In an important replication that sheds light on the gene expression patterns underlying learning, singing-induced gene coexpression patterns within Area X preserve remarkably well across the critical period boundary. The adult-preserved singing transcriptome and the adult non-preserved learning transcriptome thus overlap only in juvenile Area X, emphasizing this brain region and time point as the nexus for investigating the gene expression and interaction fundamental to the learning process.

The final chapter of this dissertation is a manuscript describing software, VoICE that was used to generate the song data that was related to gene expression to find the coexpression modules related to learned vocalization. VoICE certainly is not the first tool for analyzing bird vocalizations but its approach to removing experimenter bias beyond determining when a vocal unit starts and stops generates results that are reliable and reproducible. Similar to the methodology used to generate gene coexpression networks in WGCNA, VoICE methodology uses the relationships between discrete units in a dataset for clustering them into biologically meaningful groups. I applied many of the network principles that are used in WGCNA are applied to vocalizations in the VoICE pipeline.

Perhaps VoICE's greatest utility is the cross-species approach it enables. The field of analysis of ultrasonic vocalization (USV) in rodents is relatively young and the methods for analyzing USVs continue to develop. This situation is akin to the analysis of bird songs before the development of SAP [68]. While those in the field of USV analysis have yet to adopt a standard for analyzing their data, demand for easy to use software solutions is high. My development of VoICE bore collaborations resulting in coauthored papers that are presented in the epilogue of Chapter 3 and the Appendix to this dissertation. Assessing the consequences of manipulations to the same gene in a vocal learner and a vocal non-learner using software that outputs similar metrics allows for parallels to be drawn that are not currently possible using other contemporary software.

In this concluding chapter, I provide commentary on the larger themes I have touched on across these studies and suggest a direction for future inquiries related to the data that I have generated.

Differential Contributions of FoxP2 Isoforms to Learned Vocal Behavior

As discussed in Chapter 2, the two isoforms of FoxP2 I examined contribute to separate and seemingly unique aspects of vocal behavior. Early *in vivo* manipulation of the full-length isoform of FoxP2 in zebra finch revealed that its knockdown throughout the sensorimotor period causes poor vocal learning as assessed by recapitulation of tutor song [34]. Subsequent work in our own lab showed that its overexpression caused a similar poor-learning phenotype [35], which I replicated in the study presented in Chapter 2. These results together suggest that the dynamic downregulation of FoxP2 concurrent with UD singing and subsequent upregulation during interspersed non-singing periods is necessary for proper vocal learning.

As with FoxP2.FL, FoxP2.10+ is dynamically downregulated with UD singing [104]. The aforementioned studies did not directly manipulate FoxP2.10+, presumably permitting the molecule's behavioral regulation throughout the sensorimotor period wherein birds are practicing their songs. I present evidence that, when FoxP2.10+ is overexpressed and the link between UD singing and the molecule's downregulation is broken, that birds learn their songs well, suggesting FoxP2.10+ has little to do with vocal learning. However, I show in Chapter 2 that birds overexpressing FoxP2.10+ are incapable of inducing variability into their songs and, in fact, their songs become increasingly *invariable* with singing.

These observations prompt questions about the role of FoxP2.FL and FoxP2.10+ on the molecular level. The current mechanism by which regulation of FoxP2 is thought to drive vocal learning in un-manipulated birds is that singing and its subsequent downregulation of FoxP2.FL and FoxP2.10+ (de)-represses FoxP2's targets, creating a state of plasticity in which the animal

explores vocal space and learns its song. This hypothesis is in line with classical theory of reinforcement learning, where broad exploration of motor space is refined through practice before arriving at the "ideal" precise pattern for execution of the skill [140,141]. This hypothesis suggests that a finch learning its song does so in a manner similar to a human learning to shoot a basketball or to swing a golf club: Motor output (song, manipulation of the ball/club) is evaluated against some internal evaluator of success (match to a tutor, outcome of the shot/swing). We then view FoxP2's downregulation and (de)-repression of its targets as necessary for driving the behavior to a variable state. Indeed, numerous lines of evidence have converged on variability as a positive predictor of learning [142-145].

These hypothesis are directly contradicted by the FoxP2.10+ birds discussed in Chapter 2, whose songs become increasingly *invariable* with singing, seemingly removing motor exploration from the equation. One would predict these animals to learn very poorly, but my results show no difference in their learning from that of control birds injected with the GFP virus. This prompts the obvious question: How do FoxP2.10+ birds learn so well without exploring motor space? Work presented in Appendix 1 wherein the modulation of variability in the songs of Bengalese finches was examined may provide a clue.

Like zebra finches, Bengalese finches downregulate FoxP2 in Area X when singing UD songs. My analyses of their vocalizations in Appendix 1 shows a case strikingly similar to the FoxP2.10+ zebra finches: Bengalese finch songs become less variable with singing [89]. This unexpected result in a manuscript written early in my graduate career provides important context for an observation made much later. Together, these observations suggest that the ability to

transition between low and high variability states is necessary for vocal learning. Our view up until my results, largely through analysis of zebra finch songs, has been that singing induces variability in the song and non-singing induces invariability. The transition between these two states is required for proper learning. The FoxP2.10+ zebra finches and Bengalese finches are the opposite: singing induces vocal invariability while non-singing induces vocal variability. In both cases, the animals are able to transition from states of low and high variability. This is not the case for FoxP2.FL birds. Both my results and those of Heston & White [35] show that FoxP2.FL overexpression removes the ability of birds to transition between low and high variability states such that periods of singing or non-singing are of equivalent variability. Among the groups that we generated, only in these animals is vocal learning affected.

In the study where FoxP2 was knocked down throughout sensoriomotor learning, the knockdown shRNA hairpins were directed against a region of the FoxP2 transcript that is not present in FoxP2.10+. The ability of these animals to modulate vocal variability was not assessed [34,146]. To understand the relationship between FoxP2 regulation, vocal variability, and learning outcome, experiments where FoxP2.FL or FoxP2.10+ are knocked down and the consequence on vocal variability is assessed are necessary. I hypothesize the knock down of FoxP2.FL would cause a similar lack of transition between low and high variability states as observed in the FoxP2.FL overexpressing birds in my work. If FoxP2.10+ were knocked down, a similar break between its expression level and behavior as in my data would occur. If a similar phenotype wherein vocal variability is decreased with singing in FoxP2.10+ knockdown zebra finches, this would make a stronger case for this isoform as a powerful regulator of vocal variability.

Molecular mechanisms of FoxP2.FL and FoxP2.10+ and their effect on behavior

The molecular function of FoxP2.FL has been clearly described. Homo- or heterodimerization between FoxP family members occurs at the zinc finger and leucine zipper domain, which precedes translocation to the nucleus and binding to DNA to affect transcription [147,148]. Significantly less is known about FoxP2.10+. Much of the hypothesis we formed prior to my work was based upon the first study in which FoxP2.10+ was directly investigated [61]. In this study, a version of FoxP2.10+ fused to an N-terminal Xpress tag was expressed in SH-SY5Y cells. Its expression pattern, in sharp contrast with that of FoxP2 isoforms containing the FOX domain, was cytoplasmic and aggregated. This result suggests that FoxP2.10+ may work as a cytoplasmic 'sink' that, due to the presence of the dimerization domain, binds other FoxP2 isoforms and prevents them from entering the nucleus. However, luciferase assays performed in HEK293T cells indicated that exogenous FoxP2.10+ expression still permitted suppression of the promoter in front of the luciferase gene, albeit in a slightly and statistically significantly weaker fashion than exogenous full-length FoxP2. Notably, the disease causing variants of FoxP2 caused far weaker suppression of the promoter than FoxP2.10+. These results are seemingly at odds: If FoxP2.10+ forms aggresomes in the cytoplasm and has no means of entering the nucleus, how does its expression cause at least partial repression of a target promoter? Unfortunately, the results obtained in this study come from SH-SY5Y and HEK293T cells, which do and do not express endogenous FoxP2. Thus, the subcellular localization and biological function of FoxP2.10+ is unclear when it is expressed in the presence of endogenous FoxP2. A parsimonious conclusion from these results is that FoxP2.10+ *modulates* but does not abolish the ability of FoxP2.FL to translocate to the nucleus and affect transcription.

In Chapter 2, I provide evidence that similar mechanisms operate in the zebra finch *in vivo*. If FoxP2.10+ were to completely inhibit endogenous FoxP2.FL availability to the nucleus, I would expect to observe a behavioral phenotype similar to that of the knockdown where song learning is poor due to FoxP2.FL being incapable of acting in the nucleus [34]. This is clearly not the case as my data show FoxP2.10+ birds learned just as well and were less variable in their learning endpoint than controls. Therefore, like the result obtained by Vernes et al. [61], expression of FoxP2.10+ in cells with endogenous FoxP2.FL does not appear to preclude the action of FoxP2.FL in the nucleus.

Our lab has piloted studies to fractionate cells overexpressing Xpress-FoxP2.10+ to: 1) validate the overexpression of the Xpress-FoxP2.10+ protein and 2) determine if this affects the subcellular distribution of FoxP2.FL. If the mechanisms proposed by Vernes et al. [61] and in my work are correct, I hypothesize that: 1) Xpress-FoxP2.10+ signal is observed only in FoxP2.10+ virus injected samples in the cytoplasmic fraction; 2) greater endogenous FoxP2.FL signal is observed in FoxP2.10+ virus injected cytoplasmic fractions vs. control cytoplasmic fractions; and 3) greater FoxP2.FL signal is observed in control nuclear fractions vs. FoxP2.10+ virus injected nuclear fractions. Initial results indicate that the anti-Xpress tag antibody binds a protein endogenous to zebra finch striatum, thus making results difficult to interpret. Troubleshooting of this antibody is required before these experiments can be conducted and interprete.

I also show that overexpression of FoxP2.10+ *in vivo* forms a similar aggregate staining pattern to that observed by Vernes et al. [61]. My results here were obtained using an Xpress tagged form of FoxP2.10+ so as to distinguish it from the endogenous form using an anti-Xpress

antibody. While the aggregate staining pattern is consistent across studies, we did not further investigate whether these aggregates are positive for aggresomal markers (ubiquitin, γ-tubulin). This aggregate staining pattern is specific to FoxP2.10+ as it is not present following overexpression of GFP or FoxP2.FL. The preliminary immunostaining experiments presented in Chapter 2 indicate FoxP2.FL and FoxP2.10+ may colocalize at the aggregates, but additional visualization experiments are required to form a strong conclusion. We also do not include a nuclear stain in these images, making it difficult to assess whether FoxP2.FL is limited from entering the nucleus. My anecdotal impression from imaging cells in which FoxP2.10+ is overexpressed is that FoxP2.FL is difficult to detect. If, indeed, FoxP2.10+ overexpression causes the degradation of FoxP2.FL while still allowing for proper vocal learning, significant work is required to uncover how this is possible.

Dynamic Regulation of FoxP2.FL and FoxP2.10+

Further experiments are necessary to understand the consequence of FoxP2.10+ overexpression: It *does not* affect learning but it *does* affect the modulation of variability. In this section, I explore how this may be possible and suggest experiments that directly address my ideas.

The experimental design used here created birds with behaviorally-regulated levels of FoxP2.FL or FoxP2.10+ while levels of the overexpressed isoform were not behaviorally-regulated and thus were at a constant high level, in line with our qRT-PCR results. It is clear that breaking the behavioral regulation of FoxP2.FL creates a poor learning phenotype while the same alteration to FoxP2.10+ does not. The information we have pertaining to FoxP2.10+ protein is that its overexpression does not preclude endogenous FoxP2 DNA binding activity, at least at the SV40

promoter [61]. Thus, we could reasonably expect that, in the FoxP2.10+ overexpressing birds presented in Chapter 2, FoxP2.FL is regulated by behavior. Still, a striking variability phenotype exists: The FoxP2.10+ birds' songs become increasingly invariable with UD song, opposite what has been observed in control birds from multiple studies from our lab, including my own [35,66]. Overexpression of FoxP2.FL creates a phenotype where UD singing does not add variability to song [35]. How, then, does FoxP2.10+ overexpression cause a more extreme phenotype where singing not only does not add variability to song but instead adds *invariability*?

The relationship between FoxP2.FL and variability has been posited to occur through interaction with suggested target genes for D1R dopamine receptors and the intracellular signaling molecule, DARPP-32 [14,149]. D1R activation increases excitability in those medium spiny neurons (MSNs) on which it is expressed [150]. Expression of both D1R and DARPP-32 decreased following knockdown of FoxP2.FL. This eliminated social context dependent variability changes in singing such that directed songs became as variable as undirected ones [149]. In our experiments, excess FoxP2.FL would presumably drive higher expression of D1R, DARPP-32, putting Area X into a state of hyper-excitability, and make song less variable [151]. These results are supported by unpublished work from our laboratory that show excitatory DREADDs (designer receptors exclusively activated by designer drugs) activation in Area X causes song to become less variable and the reciprocal inhibition of Area X causes increased variability (Heston et al., 2017, *in revision*).

With the prior information in mind, a possible action of FoxP2.10+ emerges. We observed strongly decreased variability following UD singing consequent of FoxP2.10+ overexpression in

Area X, which the DREADDs experiments suggest correlate with high firing rates in Area X. Under the hypothesis that FoxP2.10+ prevents FoxP2.FL from entering the nucleus, birds overexpressing FoxP2.10+ would exist in a state where D1R and DARPP-32 are downregulated, prompting a state of low firing in Area X and high vocal variability that would be further accentuated by singing. However, this hypothesis is not in line with the results observed by Vernes et al. [61] as repression of the SV40 promoter was observed in the presence of FoxP2.10+ in cells where endogenous FoxP2.FL is also present, suggesting that FoxP2.10+ may play a *modulatory* instead of inhibitory role in its interaction with FoxP2.FL. Some evidence for this idea exists: FoxP family members can form heterodimers at the zinc-finger leucine-zipper domain [148] and the FOX domain of FoxP2 is capable of binding DNA as a monomer or as part of a dimer [152]. Therefore, overexpression of FoxP2.10+ may influence dimerization, regulation, and transcriptional activity of all FoxP family members as well as the dimeric to monomeric FoxP2 ratio. A summary of the relationship between endogenous and virus-driven FoxP2 isoforms, singing behavior, the transcriptional state, and behavioral output are presented in Figure 4-1.

Using these results as a guide for the *in vivo* work in zebra finch, a FoxP2.10+ virus-induced state of invariability through D1R and DARPP-32 upregulation could result from alteration of FoxP family member homeostasis, availability to the nucleus, and/or affinity for genome binding sites consequent of excess FoxP2.10+. Experiments to determine whether FoxP2.10+ overexpressing cells have greater levels of D1R and DARPP-32 despite behavioral downregulation of FoxP2.FL would provide evidence to support this hypothesis.

A key point in our AAV studies manipulating FoxP2 isoforms is that the viruses transfect a relatively small portion of Area X, estimated at ~25% of the cells at the injection site, of which ~95% are neurons [35]. This creates a complicated situation where, presumably, some cells cycle FoxP2 with behavior and others do not. Notably, this may accurately model the phenotype observed in the KE family, where random monoallelic expression likely occurs [153], such that a mosaic of "good" and "bad" FOXP2 cells intermingle to generate a speech and language phenotype. This fact may help contextualize the results we observe in Chapter 2, where virus transfected cells are interspersed among a majority of non-transfected cells.

No clear signature of FoxP2.FL overexpression is evident in the RNA-seq data presented in Chapter 2, though it was observed in the same samples using qRT-PCR. At the onset of the study, we hypothesized that we would see suppression/activation of FoxP2 target genes in the animals overexpressing FoxP2.FL. This most likely is due to a poor signal to noise ratio: the virus transfected cells are much less abundant than the non-transfected ones in the tissue punch, indicating that the gene expression profile observed is much more representative of non-transfected cells. Further, we also expected to observe a distinctive gene expression profile consequent of FoxP2.10+ overexpression such that the clear behavioral phenotype would be linked to specific FoxP2 targets whose expression is differentially modulated by the protein's overabundance. These results would provide better insight to the molecular basis of the variability phenotype resultant of FoxP2.10+ overexpression.

Single cell sequencing technologies are emerging and would remedy the signal to noise issue discussed above. One could feasibly overexpress any isoform of FoxP2, then disaggregate

the cells and sequence on the single cell level wherein the amount of FoxP2 within the cell would provide a mechanism by which to bioinformatically distinguish transfected from non-transfected cells. Differential expression and/or network analysis would then reveal the effect of FoxP2 overexpression on the transcriptome without confounding data from untransfected MSNs or other neurons and glia in the punch.

MAPK11, ATF2, HISTH2B and other building blocks of learned vocalization

A significant pathway highlighted in the discussion section in Chapter 2 is a protein-level interaction between a potential FoxP2 target (MAPK11), a network hub gene in a learning module (ATF2), and a singing related gene (HISTH2B). FoxP2 is thought to repress MAPK11, which in turn lowers the likelihood of MAPK11 phosphorylating ATF2, which then decreases ATF2 acetylation of HISTH2B. The association between these molecules was discovered through a search of the literature and subsequent experiments in the lab are underway to determine if overexpression of FoxP2/poor vocal learning is correlated to this pathway. We have obtained antibodies against MAPK11 and a phospho-specific antibody for the MAPK11 phosphorylation sites on ATF2. We expect less MAPK11 and lower levels of p-ATF2 in tissue overexpressing FoxP2.FL relative to a control, uninjected hemisphere from each animal. Should our results suggest this is the case, the next step in the pathway is to examine whether HISTH2B receives less acetylation due to diminished ATF2 activity. Ultimately, the link between acetylation of HISTH2B results in changing the transcriptional state of nearby genes, which presumably are transcribed then translated to serve a biological function. The specific genes affected by this epigenetic

modification then become of interest. Experiments where ATF2 is made active to drive HISTH2B acetylation either *in vitro* or *in vivo* would be necessary.

To address another hypothesis made in the discussion section of Chapter 2, we expect this signaling pathway to be of greater importance in 65d birds, in which ATF2 is a very well connected module and network member, than in adults where ATF2 is poorly connected. By using protein interaction databases, I suggest hundreds of protein level interactions analogous to ATF2 and HISTH2B, where a well-connected member of a learning module interacts with a singing related gene. The advantage of using networks to guide the interpretation of transcriptomic data is clear here: The prioritization of nodes by their network position allows for molecules with no *a priori* relationship to a trait to 'rise to the top' and reveal themselves as worthwhile targets for investigation.

seqFISH

Our laboratory has recently been contacted with an opportunity to collaborate using seqFISH [154], a technique by which one can visualize hundreds of transcripts across all cells in one tissue section simultaneously. This technology can be used to validate in a semi high-throughput fashion the coexpression network data I have generated. Further, this technique will fill in a key gap from the network analysis as it offers single cell resolution so that coexpression of behaviorally relevant genes can be confirmed to occur in the same cells (or not). With the presence of cell type markers, we will be able to determine the specific coexpression patterns within the cells that drive singing and learning. With this tool in hand and tissue from juvenile and

adult birds with or without FoxP2 overexpression, we will be able to address many hypotheses including but not limited to the ones listed below:

Hypothesis	Gene Set (See Ch. 2)	Region 1	Region 2
Learning-related gene coexpression is specific to juvenile striatopallidum	Green learning module	Juvenile Area X OR juvenile striatopallidum	Adult Area X or adult striatopallidum
Singing-related gene coexpression is preserved into adulthood	Singing-related modules	Juvenile Area X	Adult Area X
Singing-related gene coexpression is specific to Area X	Singing-related modules	Juvenile Area X and/or Adult Area X	Juvenile VSP and/or Adult VSP
MAPK11 and ATF2 are close network neighbors	MAPK11 and ATF2 probes	Juvenile Area X	N/A
FoxP2 suppresses MAPK11 expression	FoxP2 and MAPK11 probes	Juvenile Area X (+ FoxP2 overexpression)	N/A

Concluding Remarks

The molecules that permit the learning of language have been elusive despite years of research as interactions between genes, proteins, and regulatory RNA elements come together to yield an intricate and complex phenotype. Our collective knowledge of how these interactions sum to the behavior has developed through the emergence of phenotypes wherein speech, language, and thus socialization are impaired. In making a link between a speech and language disorder and the FOXP2 gene, an entry point to the complicated molecular network underlying language was found. I have studied the behavioral role of this gene and its role in the molecular networks

underlying vocal learning in the classical laboratory model of vocal learning, the zebra finch, whose neuroanatomy, reliance on FoxP2, and vocal development are strikingly similar to those of humans despite their considerable evolutionary distance.

My results show that the 'core' of the juvenile transcriptome (wherein birds are actively learning their songs) in a region specifically devoted to singing, Area X, is a densely interconnected set of genes that include markers for social and communication disorders. This coexpression pattern is completely lost in the same region of the adult brain (wherein song is no longer being learned). Singing related coexpression patterns are specific to Area X and largely preserved between juvenile and adult. This presents an intriguing situation where the overlap of singing and learning transcription patterns during the critical period in which finches are capable of learning their songs drive the molecular processes underlying the ability to learn vocalizations. The hypothesis that I propose based on these observations is that the closure of the critical period in which the bird can learn its song, which coincides with sexual maturity, correlates with the altered coexpression pattern of learning related genes.

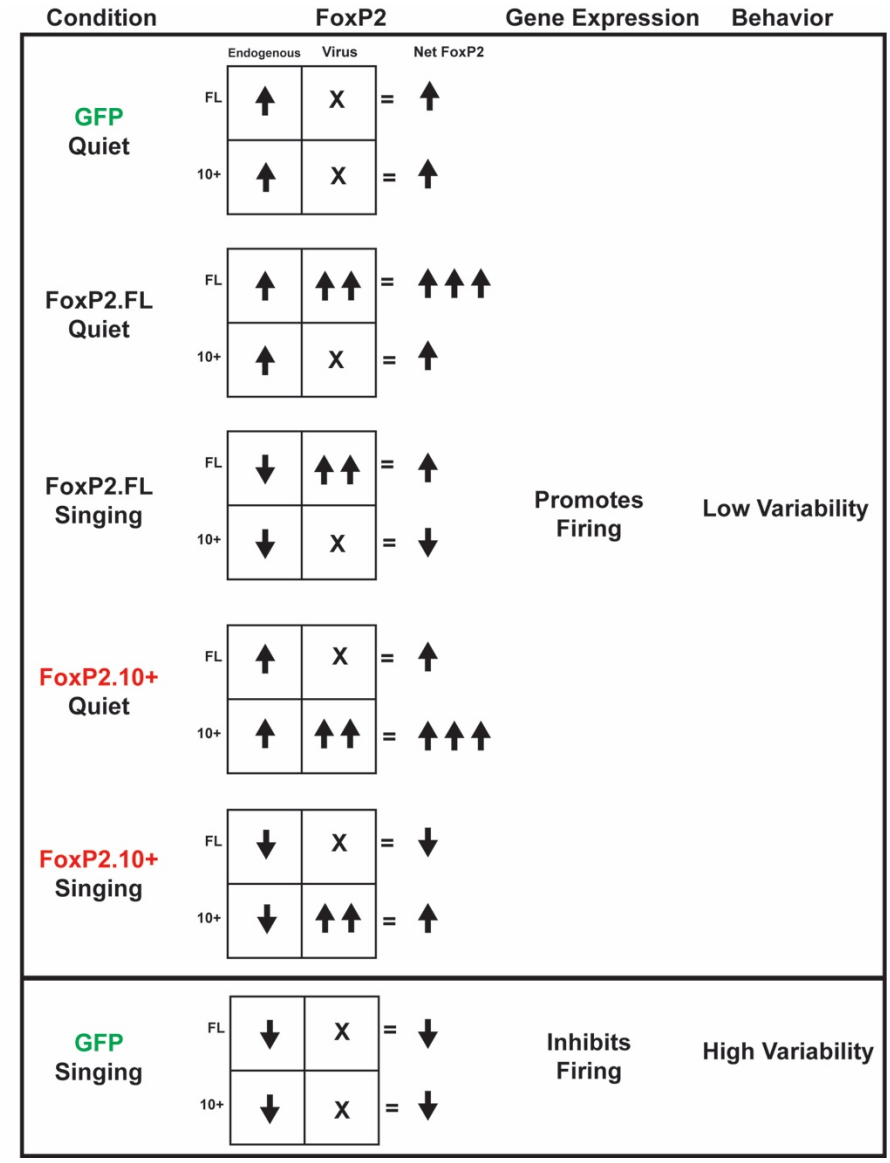
Area X is but one component of the song circuit. Like Area X, the other song nuclei are thought to have evolved from less specialized motor circuitry [10,79-81]. Further study to see if a similar principle, where a generalized plasticity related coexpression pattern nonspecific to but including the song nucleus, whether it be the same genes described in the learning related modules of Chapter 2, overlaps with song nucleus-specific singing-induced coexpression patterns is warranted. Studies in birds with seasonally plastic songs, such as canaries, would be informative. Based on my work, one would predict that canaries would have this learning related coexpression

pattern during the breeding/plastic season and not during the nonbreeding/stereotyped season. If these coexpression patterns change with the seasons, understanding the driver for this change would become exceedingly interesting as a tool by which plasticity in vocalizations can be induced.

While birdsong is not language, both behaviors require vocal learning and this unifying subcomponent has been the focus of my research. While humans are capable of adding to their vocabulary throughout life, the capacity for learning language is much greater earlier in life [55] and this observation is the basis for the critical period hypothesis of language learning. If, as the analogy between human and finch brain suggests, a similar learning-related coexpression pattern exists in the human striatum that disappears following the closure of a critical period, how does one use the information obtained herein to begin remedying language disorders? There are many hundreds of genes in the learning-related modules in the network presented in Chapter 2. In order to reopen a critical period and induce a state by which an animal can effectively learn new vocalizations, manipulation of seemingly hundreds of genes simultaneously is necessary. This approach is not feasible in the zebra finch and impossible in humans. Similarly to the proposed mechanism for inducing vocal plasticity in seasonal song learners, understanding the mechanisms by which the patterns of coexpression within this module change becomes the focus. As WGCNA has repeatedly shown its predictive power in identifying molecules relevant to a phenotype, I suggest starting with the most well connected genes in the juvenile network that are correlated to learning and using their network position as a basis for their investigation. As hub genes drive the transcriptome, so should they drive the behavior.

Figures

Figure 4-1: Summary of FoxP2 Level and Behavioral Output



Across all virus constructs and behavioral conditions, the only situation where the bird is able to induce variability into its song is when both FoxP2.FL and FoxP2.10+ are downregulated together, exemplified by the lowest panel. As discovered in our laboratory, the firing state of Area X

negatively correlates with the vocal variability state of the animal (Heston et al., 2017, *in revision*). Based on these observations together with those of Vernes et al. [62], I predict the overexpression of FoxP2.10+ modulates the transcriptional state of MSNs so as to induce a state of extreme stability. The transcriptional basis for this invariability likely lies in the positively-correlated variability induction modules presented in Chapter 2.

Appendix 1: Expression Analysis of Speech-Related Genes *FoxP1* and *FoxP2* and Their Relation to Singing Behavior in Two Songbird Species.

Qianqian Chen, Zachary D. Burkett, Jonathan B. Heston, and Stephanie A. White

Statement of Contribution

The down-regulation of FoxP2 concurrent with vocal exploration in zebra finch Area X relative to the outlying non-song striatopallidum is correlated with increases in the variability of the bird's vocalizations, akin to the induction of a state of vocal plasticity. The Bengalese finch, a similar species of songbird to the zebra finch, sings a more complex song and has a greater reliance on audition for song maintenance, as evidenced by quicker song degradation following deafening. This suggests Bengalese finches may have a greater reliance than zebra finches upon the regulation FoxP2 to drive the variability inherent in their songs. To test this, I analyzed songs of Bengalese finches sung immediately following two hours of silence or two hours of signing. Surprisingly, despite observing the expected downregulation of FoxP2 concurrent with singing, my analyses indicated that vocal variability was decreased instead of increased, as in zebra finches. These results, along with those observed in Chapter 2, suggest that the ability to transition between low and high variability states is the important driving factor for properly learning to vocalize. I contributed behavioral analyses that are depicted in Figure A1-2, A1-7, and A1-8. I wrote a portion of the methods and results sections pertaining to analyses of vocal behavior.

Summary

Humans and songbirds are among the rare animal groups that exhibit socially learned vocalizations: speech and song, respectively. These vocal-learning capacities share a reliance on audition and cortico-basal ganglia circuitry, as well as neurogenetic mechanisms. Notably, the transcription factors Forkhead box proteins 1 and 2 (FoxP1, FoxP2) exhibit similar expression patterns in the cortex and basal ganglia of humans and the zebra finch species of songbird, among other brain regions. Mutations in either gene are associated with language disorders in humans. Experimental knock-down of FoxP2 in the basal ganglia song control region Area X during song development leads to imprecise copying of tutor songs. Moreover, FoxP2 levels decrease naturally within Area X when zebra finches sing. Here, we examined neural expression patterns of FoxP1 and FoxP2 mRNA in adult Bengalese finches, a songbird species whose songs exhibit greater sequence complexity and increased reliance on audition for maintaining their quality. We found that FoxP1 and FoxP2 expression in Bengalese finches is similar to that in zebra finches, including strong mRNA signals for both factors in multiple song control nuclei and enhancement of FoxP1 in these regions relative to surrounding brain tissue. As with zebra finches, when Bengalese finches sing, FoxP2 is behaviorally downregulated within basal ganglia Area X over a similar time course, and expression negatively correlates with the amount of singing. This study confirms that in multiple songbird species, FoxP1 expression highlights song control regions, and regulation of FoxP2 is associated with motor control of song.

Introduction

The importance of the FOXP subfamily of transcription factors in the brain was not clear until FOXP2 was identified as the monogenetic locus of a speech and language abnormality. Half of the members of a British pedigree, known as the KE family, suffer from a rare communication disorder. Affected members share a single mutation in FOXP2 that causes a severe impairment in the selection and sequencing of fine orofacial movements [11,60]. In addition to articulatory problems, affected individuals have profound deficits in production and comprehension of word inflections and syntactical structure [155,156]. The phenotype resulting from its mutation indicates that FOXP2 is linked to neural pathways underlying speech and language.

FOXP1 is the closest forkhead family member to FOXP2, with which it shares high similarity at the amino acid level (68% identity and 80% similarity between the two human sequences). FOXP1 can heterodimerize with FOXP2 and can repress transcription of similar groups of genes [147,148,157]. FOXP1 is also associated with speech and language through multiple cases [158-160]. For example, a patient with a genetic deletion restricted to FOXP1 exhibits difficulties with verbal expression resembling the phenotype of affected KE family members [161]. Besides humans (Homo sapiens), no taxon of primates is capable of substantially modifying its vocal repertoire in response to experience. Moreover, most laboratory animals, including rodents, do not learn a substantial portion of their vocalizations [58,106,162]. In striking contrast, thousands of songbird species share the trait of vocal learning with humans, enabling comparison of brain–behavior relationships among these taxa. Zebra finches (*Taeniopygia guttata*) are a well-studied songbird species in which song learning is sexually dimorphic: juvenile males

learn their courtship songs from adult male conspecifics (tutors) whereas females do not produce learned songs. Zebra finch song is composed of notes, syllables, motifs and bouts. Notes are the smallest unit of song and are defined as a region of a syllable that maintains a temporally continuous frequency pattern. Syllables are composed of one or more notes bounded by a brief period of silence. Motifs are repeated sequences of syllables lasting ~1 s with multiple motifs in succession organized in a bout. Bouts are composed of several motifs bounded by a longer period of silence [108,163].

Male, but not female, zebra finches possess the full and interconnected suite of cortico-basal ganglia nuclei that underlies song learning and production. Song control circuitry includes the anterior forebrain pathway (AFP), which is important for song learning in juveniles and song maintenance and plasticity in adults, and the posterior descending pathway, which is required for song production [6,164,165]. Neurons in the HVC (acronym used as a proper name), a premotor vocal control nucleus, directly project to the robust nucleus of the arcopallium (RA) [166,167]and indirectly project to the RA through basal ganglia nucleus Area X, the medial nucleus of the dorsolateral thalamus, and the lateral magnocellular nucleus of anterior nidopallium (LMAN) in the AFP. The AFP is homologous to basal ganglia-thalamo-cortical circuit loops in mammals. Area X shares many features characteristic of the mammalian striatum and pallidum, including cell types and connectivity[168].

Songbirds and humans also share neurogenetic mechanisms that underlie their vocal learning capacities. FoxP1 and FoxP2 exhibit similar expression patterns in the cortex and basal ganglia of humans and zebra finches [31]. Knock-down of FoxP2 in Area X of juvenile zebra

finches leads to imprecise copying of the tutor song, suggesting that FoxP2 is involved in the normal process of vocal learning [34]. Moreover, Area X FoxP2 is behaviorally and socially regulated. Non-singing zebra finches have high levels of Area X FoxP2 that decline acutely when males practice their songs alone (termed undirected singing) in the morning, but not when they sing to females (directed singing)[32,51].The downregulation of FoxP2 during undirected singing is particularly robust in juvenile zebra finches undergoing sensorimotor learning: the more they practice, the lower their Area X FoxP2 levels. Interestingly, hearing is required to maintain this negative correlation [104]. Moreover, coincident with decreased FoxP2, vocal variability increases after 2 h of undirected singing in both juvenile and adult zebra finches [51,66].These observations have led us to hypothesize that singing-driven decreases in Area X FoxP2 levels promote vocal variability and motor exploration whereas high levels promote song stabilization [66].

Here, we further test the relationship between learned vocal behaviors and FoxP1 and FoxP2 gene expression by examining another songbird species, the Bengalese finch (*Lonchura striata domestica*), in which song learning and song control circuitry are also sexually dimorphic, but whose song exhibits features that are distinct from zebra finch song. Adult male zebra finches sing a linear song sequence and thus exhibit a very simple birdsong 'syntax', whereas male Bengalese finches generate songs with greater syntactical complexity [169]. After deafening, the songs of Bengalese finches degrade faster than those of zebra finches [170,171], indicating a greater reliance on audition for their song maintenance. These observations suggest that singing-driven decreases in Area X FoxP2 levels might be more robust in Bengalese finches than in zebra finches. As a consequence, increases in song variability following song practice might be evident in adult male Bengalese finches.

We therefore tested the following hypotheses: (1) Bengalese finches and zebra finches share similar FoxP1 and FoxP2 gene expression patterns; (2) FoxP2 mRNA is behaviorally regulated in male Bengalese finches; (3) downregulation of FoxP2 within Area X is correlated with the amount of undirected singing in both species; (4) the singing-driven regulation of FoxP2 within Area X of Bengalese finches is more profound than in zebra finches; and (5) vocal practice promotes song variability in adult male Bengalese finches.

Materials and Methods

We conducted in situ hybridization on brain tissue from Bengalese finches and zebra finches of both sexes under different behavioral conditions to investigate FoxP gene expression patterns, the time course of downregulation of FoxP2, and the relationship between amount of singing and FoxP2 levels within Area X. A separate group of adult male Bengalese finches was used to investigate song variability following two different behavioral conditions known to alter Area X FoxP2 levels (Figure A1-1).

Animals and tissues

All animal use was in accordance with National Institutes of Health guidelines for experiments involving vertebrate animals and approved by the Institutional Animal Care and Use Committee of the University of California, Los Angeles. Adult male and female zebra finches and Bengalese finches (age >120 days) were taken from our breeding colony (where they were kept under a 13 h:11 h light:dark cycle). After behavioral monitoring (see below), birds were decapitated for collection of brains, which were rapidly extracted and frozen on aluminum floats on liquid nitrogen, then stored at −80°C until use.

Riboprobe preparation and in situ hybridization analysis

FoxP genes are highly conserved among such disparate avian species as zebra finches and chickens (FoxP1: zebra finch versus chicken, identities=95%; FoxP2: zebra finch versus chicken, identities=97%). Although the genome of Bengalese finches is not yet available, the similarity of

their FoxP genes to zebra finch sequences is expected to be even higher based on their closer phylogenetic relationship. We therefore used riboprobes directed against zebra finch FoxP1 and FoxP2 [31] to detect these transcripts in both species. The FoxP1 probe was designed to hybridize to the coding region upstream of the zinc finger domain of zebra finch FoxP1, corresponding to 661–998 bp of human FOXP1 relative to the start codon. The FoxP2 probe was designed to hybridize to 1870–2127 bp of the zebra finch FoxP2 relative to the start codon. pCR4-TOPO vector (Invitrogen, Carlsbad, CA, USA) with zebra finch FoxP cDNA fragments was used for in vitro transcription to generate sense and antisense RNA probes labeled with [33P]UTP (Perkin-Elmer, Foster City, CA, USA) using the Riboprobe Combination System-T3/T7 (Promega, Madison, WI, USA).

Frozen brains were cryosectioned in either the sagittal or coronal plane at 20 μm and adjacent sections were mounted onto 25×75 mm slides (Superfrost, Fisher Scientific, Pittsburgh, PA, USA) in a manner that created seven replicate sets. One set was stained with thionin to enable identification of neuroanatomical structures. The adjacent four sets were exposed to the FoxP1 sense, FoxP1 antisense, FoxP2 sense or FoxP2 antisense probes. In situ hybridizations were performed and signals from different brain regions were quantified as previously described [31,32,104]. Sections of Bengalese finches were run aligned with sections of zebra finches from the same behavioral conditions to enable direct comparisons. Preliminary analysis of Bengalese finch sections revealed that: (1) the distinct expression patterns between brains exposed to either FoxP1 or FoxP2 antisense probes were as expected based on prior studies, (2) signals from antisense probes were robust whereas those from sense probes were negligible, and (3) signals were consistent across adjacent brain sections. These results provide confidence that riboprobes

designed from zebra finch cDNA also specifically detect FoxP1 and FoxP2 in Bengalese finch brain.

Behavioral monitoring and sound recording

Birds were housed individually in sound attenuation chambers (Acoustic Systems, Austin, TX, USA) for 2–3 days prior to the behavioral experiments to enable acclimation to the recording environment. Sounds were recorded using Countryman EMW omnidirectional lavalier microphones (Countryman Associates, Menlo Park, CA, USA) and digitized using a PreSonus Firepod (44.1 kHz sampling rate, 24 bit depth; Baton Rouge, LA, USA). Recordings were acquired using Sound Analysis Pro (SAP) 2011 software [68].

Behavioral experiments were conducted between 08:00 and 11:00 h, starting at lights on. For FoxP gene analysis, birds were killed following the completion of different behavioral paradigms, which are illustrated in Figure A1-1 and described as follows. Female birds were left alone and undisturbed inside the chamber for 2 h after lights on. Non-singing males (referred to as NS; Figure A1-1A) were also left alone for 2 h after lights on, but with the door to the chamber ajar. If they appeared to attempt to sing, they were distracted by the experimenter. Those that sang more than five motifs despite the experimenter's presence were excluded from this group. Of note, we previously found that the non-singing paradigm did not lead to detectable changes in zebra finch stress levels as measured by serum corticosterone values [172]. In addition, Area X gene expression patterns from birds that were distracted from singing by an experimenter clustered together with patterns from birds that sang very little by their own volition. This suggests that singing behavior – and not the absence or presence of the experimenter – is the more crucial

determinant of gene expression in Area X [51]. Males singing undirected song (referred to as UD; Figure A1-1A) were allowed to sing alone inside the chamber for a pre-determined period of time – 1, 1.5 or 2 h after the first song in the morning. For analysis of song variability, a separate set of birds was used for which the behavioral conditions are illustrated in Figure A1-1B. One group of male birds (N=6) was kept from singing for 2 h and then allowed to sing undirected songs. Songs sung during the subsequent 20 min (termed NS-UD songs) were analyzed. On another day, the same group of male birds was allowed to sing undisturbed for 2 h, and then songs that were sung in the subsequent 20 min (termed UD-UD) were analyzed.

Quantification of the amount of singing

Audio files generated by SAP were edited with Audacity 1.3 Beta (http://audacity.sourceforge.net) by manual removal of cage noise and calls, leaving only songs. In our previous study on zebra finches, the amount of singing was quantified by counting the number of motifs [32]. However, there is considerable variability in phonology and macroscopic song structure both within and between the two songbird species studied here (Figure A1-2A). The greater syntactical variability in Bengalese finch song makes it challenging to identify their motifs (Figure A1-2B). Moreover, the length of the motifs varies among different Bengalese finches and between the two songbird species. To minimize error and avoid introducing bias by manually identifying song motifs, we used SAP to automatically measure the length of each song syllable. Syllables were segmented using experimenter-derived amplitude thresholds in SAP, and then run through the 'Feature Batch' module, which computes the duration of each syllable in the batch.

The total amount of singing was then defined as the sum of the durations of all syllables identified for a given time period.

Quantification of FoxP gene expression

For semi-quantitative and quantitative analyses, optical density (OD) measurements of FoxP signals were obtained from digitized images of autoradiograms using Adobe Photoshop 7.0 (Adobe Systems, San Jose, CA, USA). First, to provide a qualitative comparison of gene expression levels across brain regions, OD values from each region were calculated from multiple sagittal sections of the brains of one 2 h UD Bengalese male, one 2 h NS Bengalese male (shown in Figure A1-3) and one Bengalese female (shown in Figure A1-4). All OD values were normalized to those from a nidopallial area of the same section that did not contain any song control nuclei. Values are reported in Table A1-1. For quantitative analysis of Area X FoxP2 levels, OD values from within Area X were normalized to those from the ventral striato-pallidum (VSP), as previously described [32].To determine the statistical significance of the Area X FoxP2 levels, a resampling procedure was employed as follows: 10,000 hypothetical data sets of the same size were resampled from the actual normalized OD values and the amount of singing in the experiments. For each resampled data set, a slope of the linear regression of these variables (OD versus amount of singing) was calculated, generating a distribution of 10,000 slopes for each species. A correlation was determined to be significantly negative if the upper and lower boundaries of the 95% confidence interval for the distribution of slopes were negative.

Syllable identification and clustering

All syllable clustering and sequence analysis was performed in the freely available R programming language (http://www.r-project.org) using custom-written clustering and syntax entropy scripts, available at the White laboratory website (https://www.ibp.ucla.edu/research/white/code.html).

To group syllables in an unbiased fashion and subsequently calculate syntax entropy, a hierarchical clustering and automated tree-trimming algorithm was utilized. Raw acoustic recordings from the first 20 min following NS or UD for each bird were subjected to SAP's 'Feature Batch', using experimenter-derived amplitude thresholds to segment syllables. A number of filtration steps were then applied to the 'Feature Batch' output to identify song syllables from non-song sounds (wing flaps, cage noise, etc.) captured by the recording software. An initial filtration step implemented user-defined duration thresholds above and below which all sounds were removed from the data set. Next, a maximum inter-syllable interval was determined by the experimenter for all remaining prospective syllables in the data set. Syllables that fell below this inter-syllable interval were grouped into prospective motifs/bouts. A filter to remove all motifs/bouts composed of two and/or three syllables was then applied. WAV files representing each motif/bout were generated and presented to the user for visual inspection, at which point motifs consisting of calls or non-song sounds in the recordings were removed from the data set if present. Finally, individual WAV files for all remaining syllables were generated.

Individual WAV files for both behavioral sessions for each animal were run against themselves in SAP's 'Similarity Batch' module in an M × N symmetric similarity batch. Upon completion of the batch, the product of the similarity and accuracy score for each syllable–syllable

comparison was calculated and stored in a square matrix with rows and columns representing individual syllables and the elements of the matrix representing the product of the similarity and accuracy scores for a given syllable–syllable comparison. A distance matrix was then created by calculating the Euclidean distance between the product of similarity and accuracy scores for all syllable–syllable pairs. This distance matrix was used as the input to a hierarchical clustering function in the WGCNA R package[119],generating a dendrogram. Branches of the dendrogram were then pruned using the dynamic tree-trimming algorithm, also in the WGCNA R package, a novel method for detecting clusters within hierarchical trees by considering the shape of the branches when trimming them into groups[45]. Upon completion of cluster detection, each cluster was described by an 'eigensyllable', defined as the first principal component of the cluster as determined by singular value decomposition. The Pearson correlation between all module eigensyllables was then computed and clusters whose eigensyllables correlated above a user-defined threshold (in this case, 0.75) were merged, generating the final number of clusters/syllable types in each bird's song.

Final inspection of cluster homogeneity was performed by visual inspection of syllable spectrograms within each cluster. Syllables inappropriately assigned to a cluster were manually reassigned.

Syntax entropy

The syllable syntax, defined as the sequence in which the bird orders its syllables, was determined based on syllable cluster assignment in the preceding step. Syntax entropy was then calculated as described in Miller et al. [66]. A string-based approach was utilized for syntax

analysis, as motifs were often difficult to identify in Bengalese finch songs. Values for syllable syntax entropy reported are weighted entropy scores, which are adjusted for the frequency of occurrence of each syllable type when determining its contribution to overall syntactical entropy. A resampling paired t-test was utilized to assess the significance of change in syntax entropy scores between behavioral conditions for all birds as a group.

Similarity, accuracy, identity and syllable acoustic features

Upon completion of clustering, syllables within each cluster were divided into NS-UD and UD-UD groups. All syllable types that did not have at least 20 renditions sung in both behavioral contexts were removed from consideration in analysis of acoustic features. The range in the number of renditions for the remaining syllables that were analyzed was 55–762. A bootstrap one-way ANOVA was performed on similarity, accuracy and identity scores and all acoustic features within each bird to determine whether syllables were independent of one another. For all acoustic measures, the between-syllable difference P-value was less than 0.05, thus syllables were treated as independent of one another.

Resampling two-way ANOVAs were performed for each acoustic measure using syllables and behavioral condition as the two independent factors. F-statistics were generated for the actual data set and then compared with a distribution of 10,000 F-statistics calculated by resampling the original data under assumption of the null hypothesis to determine whether a syllable effect, a behavioral effect and/or an interaction between the two variables were present for each measure.

Results

FoxP1 expression in Bengalese finch brain

FoxP1 mRNA signals indicated high expression levels in the densocellular part of the hyperpallium, the mesopallium, the striato-pallidum and the dorsal thalamus in both male (Figure A1-3) and female (Figure A1-4) Bengalese finches. In the basorostral pallial nucleus (Bas) and song control nucleus LMAN, FoxP1 expression was lower than in the surrounding nidopallium region regardless of sex (Figure A1-3C, Figure A1-4). In contrast, sexually dimorphic FoxP1 expression was observed in song control nuclei HVC, RA and striato-pallidal Area X, as the signals were greater in these nuclei relative to the respective surrounding brain tissue only in male Bengalese finches (Figure A1-3C). In females, signals were similar across these sub-regions (Figure A1-4). FoxP1 did not appear to be regulated by undirected singing in male Bengalese finches. Expression patterns from sagittal sections containing multiple song control regions were broadly similar between the 2 h NS and UD groups (Figure A1-3C). A semi-quantitative summary of these observations is presented in Table A1-1. Coronal sections from a separate set of birds were used to focus on Area X and LMAN (Figure A1-5), but again, no behavioral regulation of FoxP1 was observed.

FoxP2 expression in Bengalese finch brain

FoxP2 signals were lightly and uniformly distributed in cortical areas whereas they were robust in the striato-pallidum, the dorsal thalamus and the Purkinje cell layer of the cerebellum in both male (Figure A1-3) and female (Figure A1-4) Bengalese finches. No sexual dimorphism of

FoxP2 expression was observed in any of the song control nuclei except for Area X. FoxP2 expression within Area X in female Bengalese finches was similar as that of the surrounding striato-pallidum (Figure A1-4). FoxP2 expression in Area X of male Bengalese finches has reported to be lower than the surrounding striato-pallidum [146]. However, the behavioral condition of the birds used in that experiment was not specified. Here we present evidence that FoxP2 within Area X is comparable to or slightly higher than in the surrounding striato-pallidum in 2 h NS Bengalese finches but lower than in 2 h UD Bengalese finches (Figure A3-D). A semi-quantitative summary of these observations is presented in Table A1-1.

Behavioral regulation of FoxP2 within Bengalese finch Area X

In zebra finches, FoxP2 expression levels decline specifically within Area X when males engage in 2 h of UD singing in the morning [32,51,104]. To determine whether similar singing-driven changes occur in a related songbird species with distinct song features, we examined FoxP2 expression in Area X of male Bengalese finches, in parallel with that in zebra finches, and compared levels between UD and NS conditions. To confirm the behavioral regulation of FoxP2 suggested in Figure A1-2D, additional 2 h NS and 2 h UD male Bengalese finches were killed and brain tissues were sectioned coronally to display Area X bilaterally in the same section. The additional time points of 1 h UD and 1.5 h UD groups were utilized to track the time course of downregulation of FoxP2 within Area X during singing. We found that Area X FoxP2 levels were significantly downregulated at the 1.5 h UD and 2 h UD time points for both species (Figure A1-6).

FoxP2 levels within Area X in 2 h UD Bengalese finches were significantly higher than those found in 2 h UD zebra finches ($P<0.01$). In order to interpret this difference, we measured the amount of singing in both groups. We found that zebra finches in our study sang more than Bengalese finches did (means ± s.e.m., Bengalese finch 351±53 s versus zebra finch 758±166 s, Kruskal–Wallis nonparametric test, $P=0.040$). Thus, the difference in FoxP2 levels between 2 h UD Bengalese finches and 2 h UD zebra finches could reflect the difference in the amount of singing. To explore this possibility, the relationship between the amount of singing and FoxP2 levels was further examined.

Correlation between FoxP2 levels and amount of singing

Area X FoxP2 levels were negatively correlated with the amount of singing in both zebra and Bengalese finches, as illustrated by the negative slope of the linear regression lines that were fit to the data from each species (zebra finch: $P<0.0002$; Bengalese finch: $P<0.0003$; Figure A1-7). These results indicate that the more a given bird sang, the lower its Area X FoxP2 level. There was no statistically significant difference between the slopes of the two regression lines ($P>0.05$, see below), indicating that, contrary to our prediction, Bengalese finch FoxP2 levels within Area X are not more responsive to singing than those in zebra finches.

Song variability after vocal practice

Songs that were sung by adult male Bengalese finches in the 20 min period immediately following a 2 h period of UD singing (UD-UD) were compared with those sung following 2 h of non-singing (NS-UD). One expectation is that there would be no difference between the behavioral

conditions, based on prior work in zebra finches in which a difference was only observed in juveniles [66]. The other expectation is that variability after UD-UD singing would be increased relative to the NS-UD conditions, based on the overall greater variability in Bengalese song and its strong dependence on hearing. In line with a majority of our predictions, we found that for many phonological and sequential measures of song variability there were no differences between the two conditions. However, on certain measures, a slight decrease in variability was observed in the UD-UD condition relative to the NS-UD condition, in contrast to our predictions. To describe syllable variability, we examined the average within-group similarity, accuracy and syllable identity (similarity × accuracy/100) of all syllables within a cluster analyzed as a function of behavioral condition. The variability of syllable identity ($P=0.034$; Figure A1-8) was lower in the UD-UD condition, reflecting similar trends in similarity ($P=0.080$) and accuracy ($P=0.075$). We next examined the mean coefficient of variance (CV) for all syllables within a cluster. Again contrary to our predictions, the CV was lower in the UD-UD condition for individual syllable features of pitch goodness ($P=0.0002$), Wiener entropy ($P=0.004$) and mean frequency ($P=0.017$; Table A1-2). A two-way ANOVA revealed that there was no effect of behavioral condition on the mean values for any of these features. Finally, we utilized entropy-based methods similar to those of Miller et al. [66] to measure syntax variability, investigating all syllables produced during the 20 min following each behavioral condition using a string-based analysis described in that study. The results indicate no significant difference in syntax entropy between the two behavioral conditions (average NS-UD entropy=0.185, average UD-UD entropy=0.168; $P>0.05$), similar to our prior findings in adult zebra finches.

Discussion

Sexually dimorphic expression of FoxP1 in songbirds

In line with our expectations, the brain expression patterns of FoxP1 and FoxP2 in Bengalese finches are broadly consistent with those previously described in zebra finches [31], including strong mRNA signals for both factors in multiple song control nuclei and enhancement of FoxP1 in HVC and Area X relative to surrounding brain tissue. One apparent difference between the two species was in the arcopallial song control region, the RA. FoxP1 in the RA of female zebra finches is higher relative to the surrounding brain tissue [31], but this enhancement was not prominent in coronal sections of a female Bengalese finch brain (data not shown) and was not detected in sagittal sections of another female Bengalese finch (Figure A1-4). Whether this is a true species difference is unclear because we were unable to detect RA in the Nissl-stained female Bengalese finch sections, despite its visibility in sections from male brains subjected to the same staining conditions (Figure A1-3). As previously reported in zebra finches [31], the RA of male Bengalese finches exhibited FoxP1 signals that were slightly higher than those of the surrounding arcopallium. Projection neurons of the RA synapse directly onto the motor neurons that innervate the muscles of phonation, similar to direct projections of layer V motor cortical neurons onto laryngeal motor neurons in humans, and are thought to enable the capacity for vocal learning [58,173]. In the spinal cord, FOXP1 plays a crucial role in defining the columnar identity of motor neurons at each axial position, as well as organizing motor axon projections [174]. Similarly, FoxP1 may organize the RA cortical motor neuron projection to syringeal and respiratory motor neurons in songbirds.

With regard to other telencephalic song control regions, enhanced expression of FoxP1 in the HVC and Area X in male, but not female, Bengalese finches mirrors the zebra finch expression pattern. There is no evidence for singing-driven regulation for FoxP1 expression in either adult Bengalese or zebra finch brains (Figure A1-3, Figure A1-5, Table A1-1). The sexually dimorphic expression of FoxP1 in song control areas (HVC, male RA, Area X), together with the speech and language deficits associated with its mutation in humans [158-161], suggest that FoxP1 plays a role in the formation of song circuitry dedicated to singing behavior.

The expression of FoxP1 within the LMAN and Bas in Bengalese finches is low relative to the surrounding tissue, and does not exhibit sexually dimorphic patterns or singing-driven regulation. The Bas is involved in feeding and oral-manipulative behaviors other than vocalization and does not anatomically connect to the vocal control system in songbirds [175]. Because both male and female finches engage in oral movements related to feeding behavior, it is not surprising that FoxP1 levels in the Bas are similar in both sexes. In contrast, the LMAN plays a key role only in male song learning and maintenance [165,176], yet FoxP1 mRNA expression was not sexually dimorphic in this nucleus. Further investigation may determine whether the FoxP1 protein exhibits sexual dimorphism in the LMAN, as differences between transcriptional and translational levels have been observed for other transcription factors in song control circuitry [177]. Although FOXP1 mutations in humans are accompanied by language disorders, the impact of FoxP1 on song learning and production remains to be determined. Given that we did not observe behavioral regulation of FoxP1 in either species, it seems likely that its role may be in promoting the developmental differentiation of neural structures, consistent with the general role of Fox transcription factors during embryogenesis [178].

Behavioral regulation of FoxP2 in songbirds

Unlike FoxP1, FoxP2 expression in male songbirds was not enriched in the HVC or the RA, and appeared similar to levels in the HVC and RA of female brains (Figure A1-3, Figure A1-4, Table A1-1). In Area X, FoxP2 was slightly higher or comparable to levels in the adjacent VSP in NS adult male songbirds. FoxP2 expression is enhanced in the striato-pallidum of hatchling zebra finches and increases in Area X during development [31]. This observation, together with the structural deficits in the basal ganglia of affected KE family members, is consistent with a role for FoxP2 in contributing to the structural organization of basal ganglia regions critical for vocal learning. Post-embryogenesis, Area X FoxP2 levels are downregulated after undirected singing in juvenile and adult zebra finches [32,104]. Lentiviral-mediated FoxP2 knockdown in Area X of juvenile zebra finches results in inaccurate copying of the tutor song [34]. Together, these findings suggest that FoxP2 is involved not only in forming neural structures for vocal learning during embryogenesis, but also in the ongoing use of such structures during vocal learning and adult song maintenance, including in adult male Bengalese finches.

Correlation between Area X FoxP2 levels and undirected singing in two species of songbird

We investigated the time course over which FoxP2 levels are first observed to decrease in Area X during singing in both Bengalese and zebra finches. We found that levels became significantly downregulated at the 1.5 h time point in both species (Figure A1-6). Contrary to our prediction, Area X FoxP2 downregulation in Bengalese finches was not more robust than in zebra finches. This outcome is qualified by the recognition that experimental quantification of the amount of singing is not always proportional to the time spent singing. For example, one zebra

finch sang for 241 s within 2 h, whereas another sang for 487 s within 1 h. We observed a negative correlation between the amount of singing and FoxP2 levels within Area X of zebra finches, which confirms results from our prior studies [32,51,104]. We now report a similar negative correlation in Bengalese finches (Figure A1-7B). Thus, singing may promote FoxP2 mRNA degradation, possibly through miRNA regulation [179], or inhibit mRNA synthesis following song onset. In either case, this regulation of FoxP2 takes time, only producing significant decreases 1.5 h following song onset in this study (Figure A1-6). It is difficult to disentangle the effects of time and the amount of singing on FoxP2 levels because we cannot control the amount and timing of singing once birds start. For each species, in birds that did sing similar amounts of song (Figure A1-7), there is a trend that the longer they were given before being killed, the lower their Area X FoxP2 levels.

FoxP2 downregulation within Area X in Bengalese finches and zebra finches

When all birds are considered, the downregulation of FoxP2 did not occur on a faster time scale in Bengalese finches than in zebra finches, as demonstrated by the lack of a statistically significant difference in the slopes of regression lines plotted to the data (Figure A1-7). The lack of a detectable difference between the two species may be due to a lack of sensitivity in the in situ hybridization. However, in pilot experiments, we compared FoxP2 levels obtained with quantitative reverse transcriptase PCR from cDNA obtained from unilateral punches of Area X with those obtained from in situ hybridization of the remaining hemi-sections from the same bird (J. Liu, unpublished). The sensitivity was comparable across methods, indicating the suitability of our approach, which also enables us to compare our current findings with past studies that

employed in situ analyses. The relationship between FoxP2 and singing in Bengalese finches may be underestimated here simply because they sang less as a group. A broader range of singing might enable detection of more subtle differences between the species. Alternatively, the dependence of FoxP2 levels on singing may indeed be similar in both species, despite differences in features of their songs.

Vocal variability after vocal practice

We previously found that in juvenile (75 days of age) zebra finches, vocal practice for 2 h in the morning leads to increased vocal variability [66] and that in adult zebra finches, the amount of singing is correlated with increased spectral entropy [51]. Thus, we predicted that vocal practice might lead to increased vocal variability in adult Bengalese finches. To our surprise, we found that despite similar behavioral regulation of FoxP2 in Bengalese and zebra finches, periods of low FoxP2 are associated with slight decreases in variability of multiple features in Bengalese finch song[51,66]. Thus, it is possible that FoxP2 downregulation may decrease vocal variability or that changes in FoxP2 levels are unrelated to changes in vocal variability in this species. Arguing against these possibilities is the observation that viral knockdown of FoxP2 in Area X is sufficient to increase variability in both juvenile [34] and adult zebra finches [149]. Multiple factors could contribute to the observed difference in these select song features, and are detailed below.

The amount of singing performed by each species could influence whether song is more or less variable in the UD-UD condition. Bengalese finches in our study sang roughly half as much as the zebra finches and the corresponding downregulation of FoxP2 is about half the magnitude. It is possible that FoxP2 levels must drop below a critical threshold in order to de-repress gene

transcription and initiate molecular changes that lead to increased variability, or that the amount of singing by Bengalese finches was sufficient to downregulate FoxP2 mRNA but not the protein [172]. These possibilities could be supported by examining Bengalese song after more extended bouts (~4 h) of UD singing; however, this may be confounded by the fact that FoxP2 levels vary as a function of both the amount of singing and the total time allotted for singing (Figure A1-7).

The age of the Bengalese finches used here (>300 days) may present another confounding factor in our ability to detect differences in vocal variability between NS-UD and UD-UD birds. Increased song variability was previously observed to be correlated to the amount of song in younger adult zebra finches [N=18 between 120 and 200 days old [51]]. Both Bengalese and zebra finches undergo age-related changes in vocal quality and the ability to exhibit vocal plasticity [180,181], thus they may undergo age-related changes in how molecular microcircuits impact behavior. Further, age- and species-related differences in basal vocal variability may have statistically limited our ability to detect these changes. In zebra finches, our ability to detect acute regulation of vocal variability was limited to 75-day-old juvenile birds, as 65-day-old birds and a group of six adult birds showed too much and too little variability, respectively, to derive adequate statistical power [66]. A follow-up study found that statistical power was achieved when the number of adult zebra finches was increased to 18 UD singers with higher numbers of motifs uttered in the 2 h being correlated with increased song variability (Figure A1-3B)[51].

In summary, these data indicate that FoxP1 is enriched in most song control nuclei of male Bengalese finches, with the notable exception of the LMAN, similar to its expression pattern in zebra finches. No singing-driven regulation of this transcription factor was observed in either

species, suggesting a sexually dimorphic role in the formation of brain structures that support vocal learning in songbirds. In contrast, FoxP2 levels in Area X do exhibit singing-driven decreases in both species, with a similar dependence on both the amount of singing and the time since song onset, with the caveat that Bengalese finches in our study sang less than zebra finches. The impact of this downregulation in zebra finches appears to be to increase vocal motor exploration, particularly during song learning and as evidenced by multiple prior studies. Here, in Bengalese finches, we did not observe a similar relationship, which could reflect a true species difference. We deem it more likely that the differences in age and amount of singing of the Bengalese finches in our study relative to the zebra finches precluded detection of this relationship. Future work in songbirds to examine protein expression of these factors as well as to genetically intervene in their expression promise to illuminate organizational versus activational functions of these molecules related to human language.

Acknowledgements

Lily Sung assisted in the preparation of brain sections. The authors thank Dr Julie E. Miller and two anonymous reviewers for their constructive comments on the manuscript.

Table A1-1: Mean optical density values measured from multiple sections of one UD male Bengalese finch, one NS male Bengalese finch and one female Bengalese finch

	FoxP1			*FoxP2*		
Brain region	UD male	NS male	Female	UD male	NS male	Female
Arcopallium	0.43	0.57	0.46	0.57	0.56	0.80
Area X	1.41	1.31	N/A	1.94	2.70	N/A
Basorostral pallial nucleus	0.62	0.68	0.54	0.72	0.89	1.02
Dorsal thalamus	1.05	1.17	1.02	3.47	4.19	4.41
Hyperpallium, apical	0.91	0.99	0.95	1.01	0.97	0.89
Hyperpallium, densocellular	1.45	1.40	1.22	1.19	1.12	1.21
HVC	1.31	1.24	N/A	0.99	1.03	N/A
LMAN	0.59	0.73	0.54	0.92	1.07	1.02
Mesopallium	1.45	1.40	1.23	1.24	1.20	1.29
Nidopallium	1.00	1.00	1.00	1.00	1.00	1.00
Robust nucleus of arcopallium	0.58	0.71	N/A	0.62	0.57	N/A
Striato-pallidum	1.30	1.30	1.17	2.42	2.48	2.74

All values are normalized to the mean optical density value in the nidopallium outside of song control areas.
NS, non-singing; UD, undirected song; HVC, letter-based name, located in the nidopallium; LMAN, lateral magnocellular nucleus of anterior nidopallium.

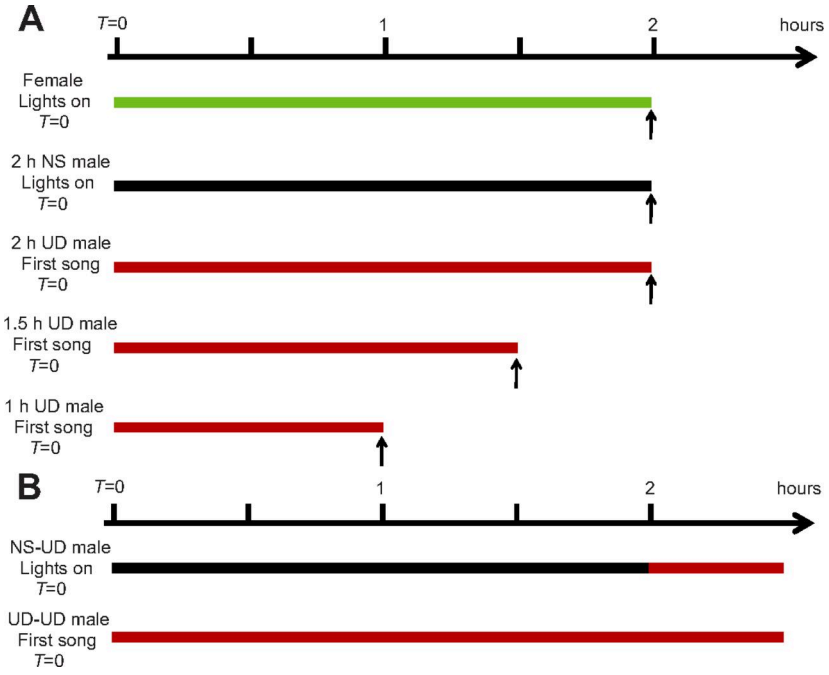
Table A1-2: Mean (±s.d.) coefficient of variation values for each acoustic feature.

Acoustic feature	NS-UD	UD-UD	*P*
Mean values			
Mean pitch	0.290±0.251	0.282±0.272	0.7169
Frequency modulation	0.147±0.068	0.136±0.058	0.1242
Entropy	0.127±0.053	0.120±0.055	**0.0376**
Pitch goodness	0.226±0.121	0.193±0.110	**0.0002**
Mean frequency	0.166±0.081	0.151±0.078	**0.0168**
Variance			
Frequency modulation	0.206±0.069	0.198±0.087	0.3800
Entropy	0.500±0.259	0.440±0.184	**0.0136**
Pitch goodness	0.662±0.319	0.581±0.412	0.1078
Mean frequency	0.895±0.678	0.775±0.328	0.1901

The mean CV for all syllable clusters within each behavioral condition is reported along with *P*-values generated by a two-tailed paired bootstrap test. Significant values are shown in bold.

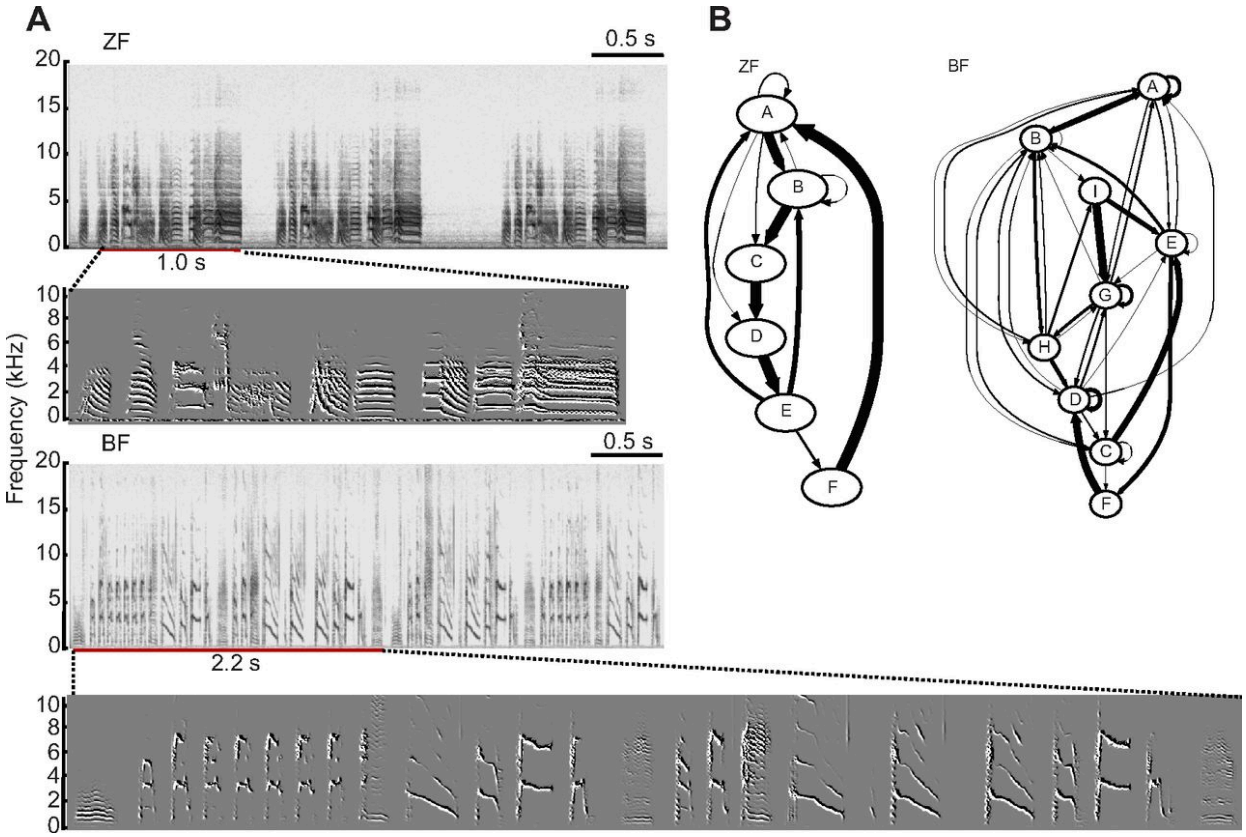
Figures

Figure A1-1: Timelines for the behavioral groups used in this study



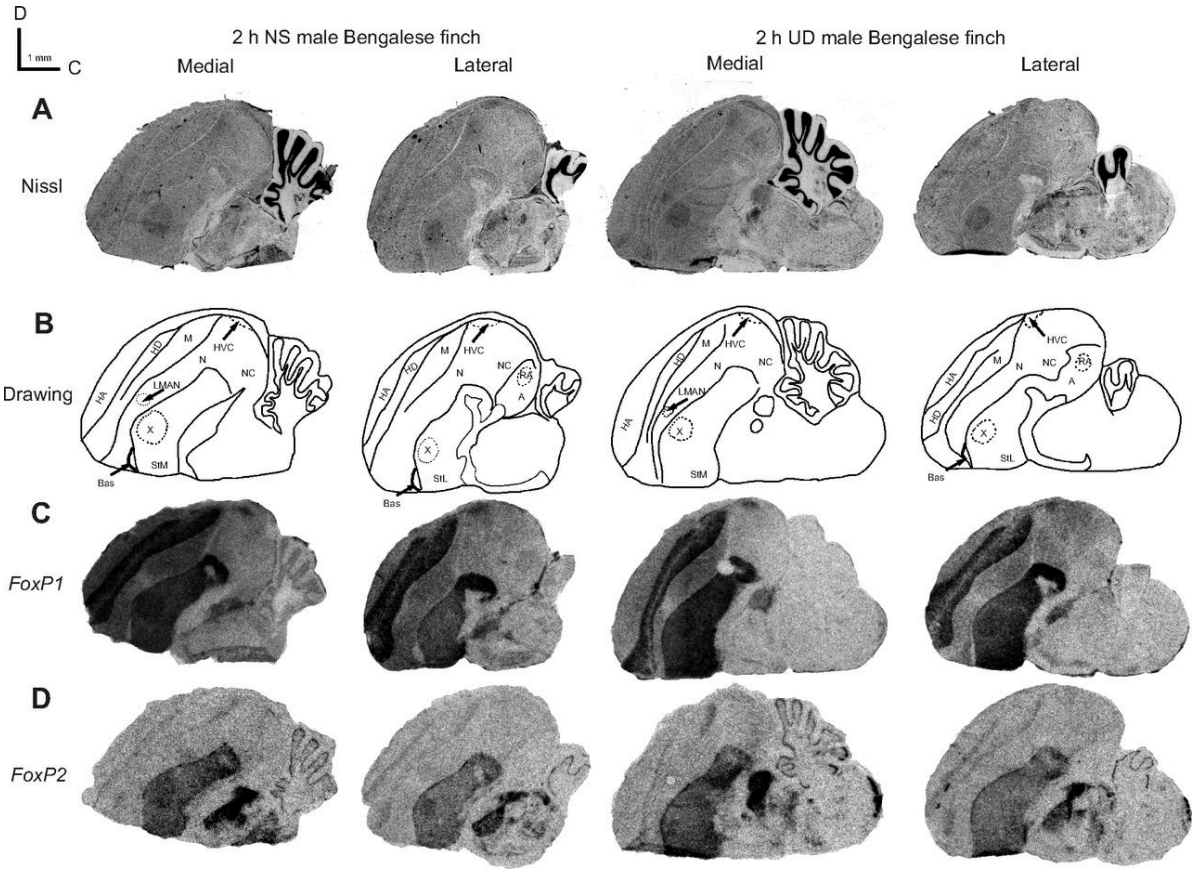
a) Experimental design for time-course analysis of FoxP1 and FoxP2 behavioral regulation. On the day of the experiment, female birds remained alone in sound attenuation chambers for 2 h (green bar). NS males were discouraged from singing by the experimenter sitting nearby for 2 h (black bar). UD males sang alone in the isolation chamber for variable periods of time (red bars). Arrows indicate the time points at which birds were killed. **b)** Experimental design for song variability analysis. Songs sung after the 2 h time point were analyzed for song variability. Birds were not killed in this experiment.

Figure A1-2: Representative exemplars of zebra and Bengalese finch song.



a) Spectrograms from a male zebra finch (ZF, top) and a male Bengalese finch (BF, middle) are shown. The red bar underneath each spectrogram indicates the length of one motif. Spectral derivatives of these motifs are shown underneath each spectrogram. **b)** Markov chains generated from zebra finch and Bengalese finch songs. Letters denote syllables. Arrows represent the probability of syllable transitions. Thicker arrows indicate greater probabilities.

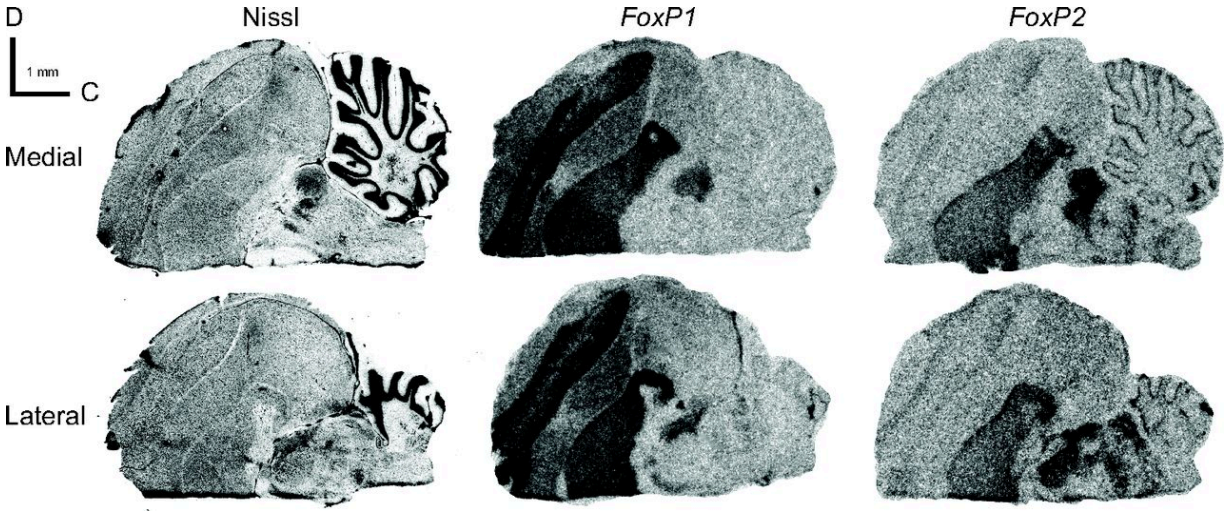
Figure A1-3: Representative brightfield photomicrographs of FoxP1 and FoxP2 mRNA expression patterns in a series of sagittal sections from one 2 h NS (left) and one 2 h UD (right) adult male Bengalese finch brain. Both medial and lateral sections are shown to enable display of the song control nuclei investigated here.



a) Nissl-stained sagittal sections. Locations of medial and lateral sections correspond to the level of sagittal plates 6 and 11, respectively, in the zebra finch atlas of Nixdorf-Bergweiler and Bischof [182]. **b)** Schematic drawings based on the Nissl stains. A, arcopallium; Bas, basorostral pallial nucleus; HA, apical part of the hyperpallium; HD, densocellular part of the hyperpallium; LMAN, lateral magnocellular nucleus of anterior nidopallium; M, mesopallium; N, nidopallium; NC, caudal nidopallium; RA, robust nucleus of arcopallium; StL, lateral striatum; StM, medial striatum. **c)** FoxP1 mRNA signals. **d)** FoxP2 mRNA signals. Medial sections in A, C and D were

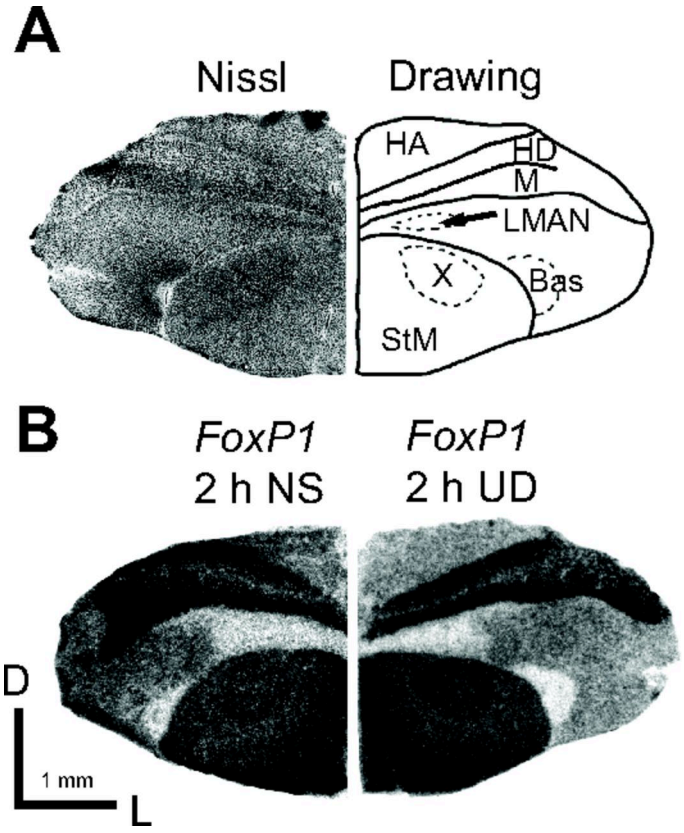
adjacent or near adjacent to one another; similarly, lateral sections were adjacent or near adjacent. D, dorsal; C, caudal.

Figure A1-4: Representative brightfield photomicrographs of FoxP1 and FoxP2 mRNA expression patterns in a pair of sagittal sections from adult female Bengalese finch brain.



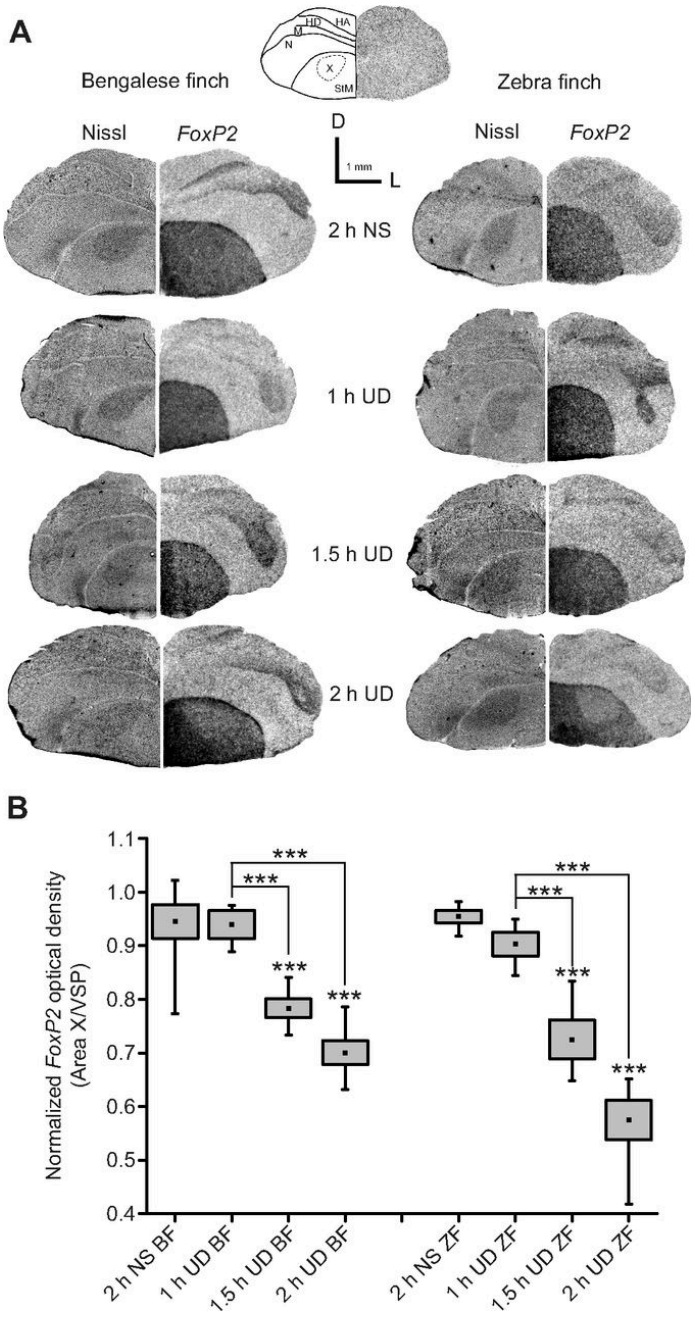
Locations of medial and lateral sections correspond to the level of sagittal plates 6 and 11, respectively, in the zebra finch atlas of Nixdorf-Bergweiler and Bischof [182]. The medial plate shows the HVC and LMAN, and the lateral plate shows the HVC and RA in corresponding sections from male birds (Figure A1-3).

Figure A1-5: *FoxP1* mRNA expression in Area X of adult male Bengalese finches.



a) Brightfield photomicrograph of Nissl-stained hemi-coronal section with schematic drawing highlights song nuclei LMAN and Area X. Abbreviations as in Fig. 3B. **b)** Representative images of FoxP1 mRNA expression at the level of Area X in 2 h NS (left) and 2 h UD (right) adult male Bengalese finch brain. There is no apparent effect of singing on expression levels. Location of sections corresponds to the level of transverse plate 11 in the zebra finch atlas of Nixdorf-Bergweiler and Bischof [182]. D, dorsal; L, lateral.

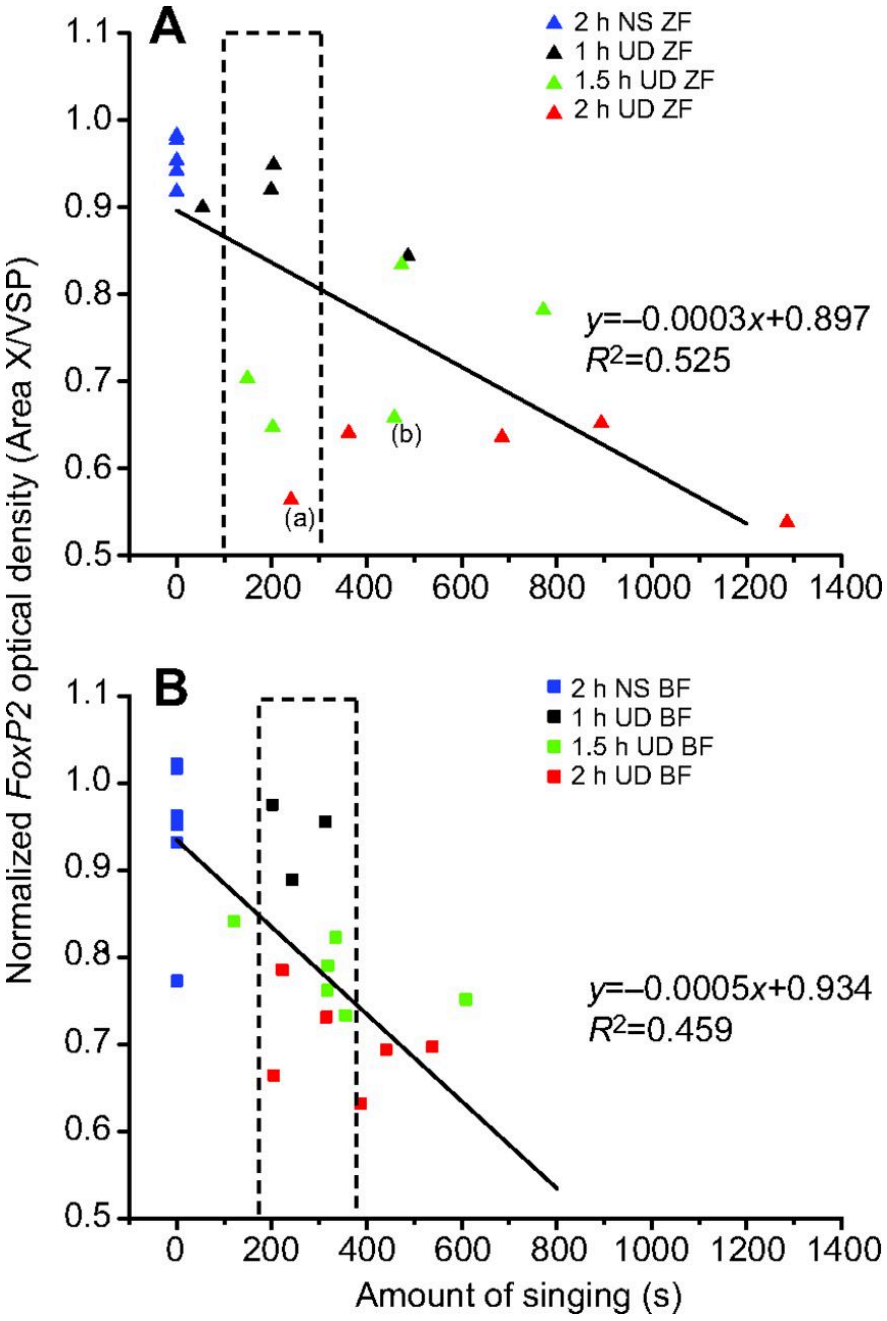
Figure A1-6: *FoxP2* mRNA expression within Area X diminishes after birds sing undirected songs.



a) Top: schematic drawing based on a Nissl-stained hemi-coronal section shown with a control hemi-section incubated with sense RNA. Abbreviations as in Figure A1-3B. Bottom:

representative brightfield photomicrographs of FoxP2 mRNA expression patterns in hemi-coronal sections from Bengalese finches (left) and zebra finches (right) of different behavioral groups shown with corresponding Nissl-stained hemi-sections. D, dorsal; L, lateral. **b)** Quantitative results of FoxP2 mRNA expression level within Area X relative to the ventral striato-pallidum (VSP). Boxes indicate s.e.m., points in boxes indicate means and whiskers indicate maximum and minimum values (2 h NS BF: N=7; 1 h UD BF: N=3; 1.5 h UD BF: N=6; 2 h UD BF: N=6; 2 h NS ZF: N=5; 1 h UD ZF: N=4; 1.5 h UD ZF: N=5; 2 h UD ZF: N=6; Kruskal–Wallis nonparametric ANOVA, ***P<0.001).

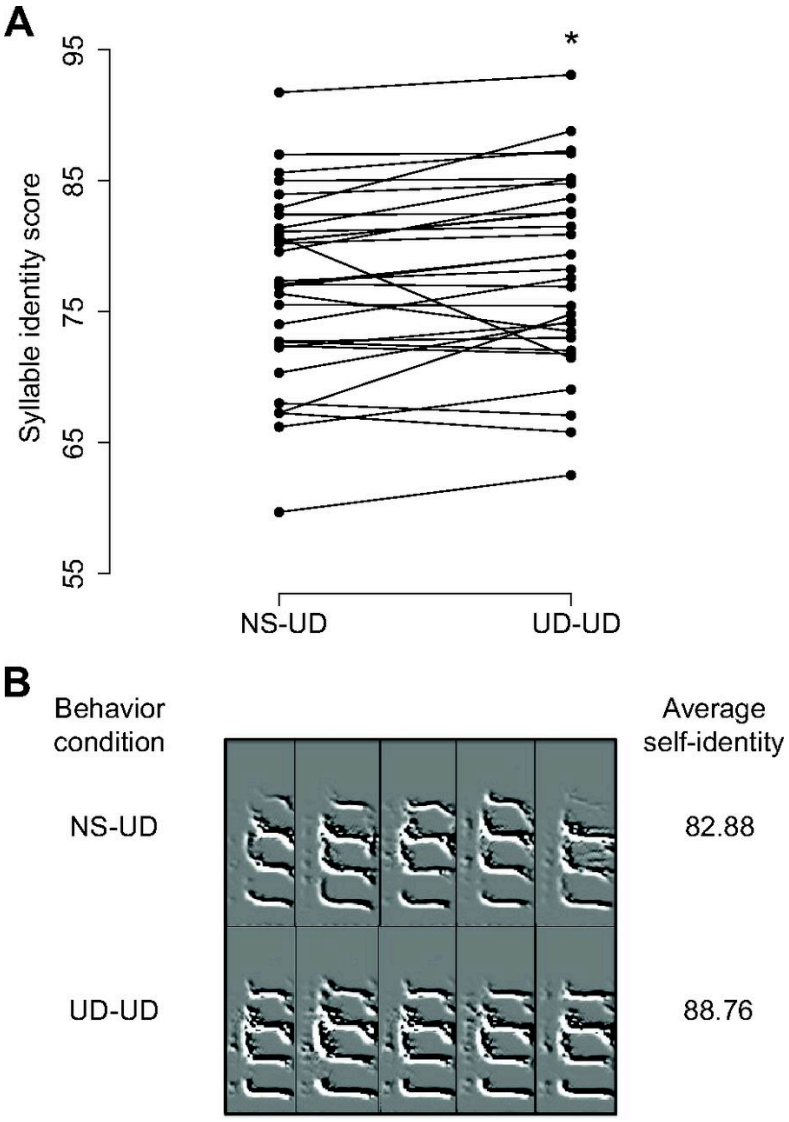
Figure A1-7: Correlation between FoxP2 and amount of singing.



a) In zebra finches, FoxP2 levels decrease as the amount of singing increases (P<0.0002). **b)** FoxP2 levels also decrease as the amount of singing increases in Bengalese finches (P<0.0003). There is

no significant difference between the two regression lines ($P>0.05$). The dotted rectangle indicates data from those birds that sang similar amounts of song for each species (see Discussion).

Figure A1-8: Behavioral changes in syllable self-identity.



a) Paired plot of syllable cluster self-identity in NS-UD and UD-UD conditions. The UD-UD condition had higher mean self-identity (*P=0.034, two-tailed paired bootstrap). **b)** Representative spectral derivatives of five syllables from one cluster in the NS-UD and UD-UD conditions with self-identity scores reported.

Appendix 2: Reduced Vocal Variability in a Zebra Finch Model of Dopamine Depletion: Implications for Parkinson's Disease.

Julie E. Miller, George W. Hafzalla, Zachary D. Burkett, Cynthia M. Fox, Stephanie A. White

Statement of Contribution

Both dopamine and FoxP2 modulate vocal variability in Area X. When a zebra finch sings an undirected song, suites of genes enriched for Parkinson's disease markers are up are concurrently up- or downregulated within Area X. Interference with FoxP2 affects dopamine signaling pathways and alters vocal variability but the precise role of dopamine in learned vocal behavior at the time of this manuscript's publication was undescribed. The work presented here is that which I pioneered as part of my M.S. degree and followed upon during my Ph.D. My contributions were in piloting the behavioral paradigms, surgical techniques, and behavior analyses. My work is presented in Figure A2-4, Figure A2-5, Figure A2-6, Figure A2-7, which I generated. I also wrote the portions of the methods and results sections that pertain to behavior and statistical analyses.

Abstract

Midbrain dopamine (DA) modulates the activity of basal ganglia circuitry important for motor control in a variety of species. In songbirds, DA underlies motivational behavior including reproductive drive and is implicated as a gatekeeper for neural activity governing vocal variability. In the zebra finch, *Taeniopygia guttata*, DA levels increase in Area X, a song-dedicated subregion of the basal ganglia, when a male bird sings his courtship song to a female (female-directed; FD). Levels remain stable when he sings a less stereotyped version that is not directed toward a conspecific (undirected; UD). Here, we used a mild dose of the neurotoxin 6-hydroxydopamine (6-OHDA) to reduce presynaptic DA input to Area X and characterized the effects on FD and UD behaviors. Immunoblots were used to quantify levels of tyrosine hydroxylase (TH) as a biomarker for DA afferent loss in vehicle- and 6-OHDA-injected birds. Following 6-OHDA administration, TH signals were lower in Area X but not in an adjacent subregion, ventral striatal-pallidum (VSP). A postsynaptic marker of DA signaling was unchanged in both regions. These observations suggest that effects were specific to presynaptic afferents of vocal basal ganglia. Concurrently, vocal variability was reduced during UD but not FD song. Similar decreases in vocal variability are observed in patients with Parkinson disease (PD), but the link to DA loss is not well-understood. The 6-OHDA songbird model offers a unique opportunity to further examine how DA loss in cortico-basal ganglia pathways affects vocal control.

Introduction

Songbirds offer attractive models for investigation of brain–behavior relationships at gene, circuit, and organismal levels. They share similar reciprocally connected cortico-striatal loops with mammals, but offer the additional advantage of a well-characterized neural circuitry for vocalization. Critically, the loci for song production are neuroanatomically distinct and well-characterized including Area X, the specialized subregion of the songbird basal ganglia dedicated to vocal learning and maintenance (Figure A2-1A). As in mammals, the basal ganglia receive dopaminergic innervation from the midbrain ventral tegmental area (VTA) and substantia nigra pars compacta (SNc) [183,184]. Feedback loops exist between these regions and Area X, as in mammalian striatum [185,186].

Dopamine (DA) orchestrates a delicate balance within mammalian and songbird basal ganglia circuits in processes associated with motor exploration versus performance and reward-based behavior [187]. The songbird model enables further exploration of the role of DA in these processes, given that it contains medium spiny neurons (MSNs) and globus-pallidal neurons in the basal ganglia, both of which are DA-sensitive [188]. Both cell types share similar anatomical and physiological signatures with their mammalian counterparts [8,189]. Levels of DA in the songbird basal ganglia are closely associated with social and breeding contexts. For example, in Area X of male European starlings and zebra finches, DA levels increase in the presence of a conspecific female, underlying his motivation to sing, as assessed through measurements of the rate-limiting catecholamine biosynthetic enzyme, tyrosine hydroxylase (TH) [190] and microdialysis of DA metabolites [191]. In contrast to the female-directed (FD) song, DA levels are lower during

undirected (UD) song, when the male sings alone [192]. DA cells in the VTA provide one source of neuromodulation onto MSNs to regulate these social-context-dependent singing behaviors [193]. Targeting these Area X inputs using neurotoxins can yield insight into the neuromodulation of vocal behavior.

In rodents, injecting the neurotoxin 6-hydroxydopamine (6-OHDA) into either the medial forebrain bundle or striatum poisons DA nerve terminals measured by TH immunostaining. This results in motor phenotypes characteristic of Parkinson disease (PD) with altered ultrasonic vocalizations detected at 72 h and 4 weeks post-injection [194]. In zebra finches, a unilateral injection of 6-OHDA into the VTA/SNc reduces TH immunostaining in Area X [192]. Consequently, FD song becomes slower. No overt changes in motif structure were noted, but acoustic variations in individual syllables were not assessed.

Here, 6-OHDA was injected directly into Area X to assess consequences on DA biomarkers and song. We hypothesized that 6-OHDA administration to Area X would reduce DA signal during both UD and FD song and lead to changes in song features. We also predicted that UD would be more sensitive to DA depletion given that levels are relatively low under this condition [192]; further reduction could drain any reservoir of signal. In normal adult males, higher levels of DA during FD song [191] are associated with less vocal variability but when D1 receptors are blocked, FD song resembles the more variable UD song [195]. Similarly, we predicted that with experimental DA depletion following 6-OHDA injection, acoustic features of FD song would resemble UD. To measure changes in DA signal in the basal ganglia, TH and dopamine receptor-associated postsynaptic protein (DARPP-32) levels were quantified using an approach, not feasible

in rodent models, of separately micro-punching vocal versus non-vocal subregions followed by immunoblotting for DA biomarkers. Below, evidence is provided for depletion of presynaptic DA terminals specifically within vocal Area X and associated changes to UD, but not FD, song. No apparent effect of experimental DA depletion in the social context was detected.

Methods

Subjects

All animal use was approved by the Institutional Animal Care and Use Committee at the Universities of California Los Angeles and Arizona. For the experiments 25 birds were used, including a subset for tissue analyses. Adult male zebra finches (120–400 days) were moved to individual sound attenuation chambers and acclimated under a 13:11 h light:dark cycle. Behavioral experiments were conducted in the morning from lights-on until overdose with inhalation anesthetic. Brain tissue for immunoblotting and immunohistochemistry was collected from birds that were euthanized immediately following lights-on (0 h nonsinging, NS, Fig. A2-1B) to prevent any confound of behavioral context and/or circadian changes.

Behavior

Methods followed those of Miller et al. [172] with some time-course modifications (Figure A1-1B). UD song recording was ongoing. FD song was captured on the day prior to or the morning of surgery and on day 4 or 5 post-surgery, depending on when singing levels were sufficient. Non-vocal behavior was simultaneously video-taped pre and post-treatment during song recording. An experimenter blind to the treatment scored the pre versus post-surgery occurrence of non-song behaviors during 30 min of UD and FD song beginning at lights-on.

Song recording and analysis

Sounds were recorded (Shure SM58/93 microphones) and digitized (PreSonus Firepod/Audiobox: 44.1 kHz sampling rate/24 bit depth). Recordings were managed using Sound Analysis Pro (SAP) [68].

The song was hand-segmented following Miller et al. [66]. Motifs were identified as a repeated order of multiple syllables, excluding introductory notes and unlearned calls. Syllables were identified as sound envelopes that could be separated from other syllables by local minima. WAV files from 20 consecutive renditions of motifs and syllables during UD and FD songs were selected (Audacity, audacityteam.org) from a similar morning time-point and run in SAP for measures of self-accuracy and individual acoustic features (SAP manual). Syllable types known as harmonic stacks were analyzed separately for changes in fundamental frequency (FF) variability [6] using code provided by M. Brainard, UCSF (MATLAB, Mathworks, Natick, MA).

Statistics and data presentation

For the means and CVs of all syllable acoustic features and self-similarity scores, resampling one-way ANOVA indicated a significant ($P < 0.05$) within-bird syllable effect. Thus, syllables were treated as independent of each other. No appreciable increase in power was observed in any statistical test when conducted on an $n > 20$ syllables in a given behavioral condition [66]. Therefore, all analyses were conducted on the first 20 syllable renditions/session.

Song features are presented in Figures A2-4–A2-7C as effect sizes. Effect size was calculated using the formula $(A - B)/(A + B)$, where A and B are measures of a given feature in two different conditions, respectively (e.g., before vs. after surgery or in FD vs. UD song). For the

effects of surgery (Figures A2-4–A2-6A), negative values indicate that a given metric is greater before injection of the drug or vehicle whereas positive values indicate the opposite. Values near zero indicate no surgical effect. For the effect of social context (Figures A2-6B and A2-7C), negative bars indicate the value is greater in UD. “Pre” denotes all pre-surgery syllables regardless of the treatment that the bird received. “Vehicle” are comparisons between post-vehicle FD and post-vehicle UD and “6-OHDA” are comparisons between post-6-OHDA FD and post-6-OHDA UD.

Effect size calculations enable statistical comparison *across* treatment groups or social contexts (i.e., 6-OHDA vs. vehicle-injected birds; UD versus FD), which is not possible using the traditional pre versus post-paired plot comparisons done *within* a treatment group. Because this calculation normalizes the scores for each measure, it is not influenced by between-syllable differences. It also allows a single number to represent the effect of a condition on a syllable feature. An unpaired resampling test on the median for each measure evaluated one condition's effect size against another's and is reported in the text. A significant effect of 6-OHDA injection was assessed when a P-value <0.05 was attained for the 6-OHDA group and not in the vehicle group.

For DA biomarkers, a resampling unpaired difference test was used to detect group differences and confirmed by a Mann–Whitney U test. Resampling was also used to assess the treatment effect (vehicle vs. 6-OHDA) on non-song behavior. For a complete description of the use of the resampling method for birdsong analysis and related citations, refer to Miller et al. [66] and Burkett et al. [69].

Surgical procedure and drug dosage

Surgery was conducted on isoflurane-anesthetized birds (n = 11, bilaterally-injected vehicle birds; n = 9, 6-OHDA birds; n = 3, received unilateral injections of 6-OHDA and vehicle within the same bird). A glass pipette was fitted into a Nanoject II pressure-injector and back-filled with mineral oil then loaded with either 0.1% sodium L-ascorbate in ddH20 (Sigma #A7631, vehicle) or 6-OHDA (Sigma, lots #MKBPO832V, #MKBR6609V) dissolved in vehicle. A 1.2 μg bilateral dose of 6-OHDA reliably spared the integrity of Area X while leading to subtle effects on song (see Results). Pilot work on the optimal dose to elicit changes in song while preserving Area X integrity indicated that variability in potency across lots of 6-OHDA necessitates testing of each new lot. In pilot work, a higher dose of 4 μg resulted in the same changes in UD song {Miller:2009to} reported in this current study at the 1.2 μg dose. However, following this work, a range of 2–4 μg doses of 6-OHDA induced a lesion in Area X that was not due to electrode damage because Area X of vehicle-injected birds remained intact. Studies of rodent models of 6-OHDA have also reported these deleterious effects, using a 6–8 μg dose (M. Ciucci, pers. comm. 2010-2011). Given these considerations, each new vial of drug was tested. Lower doses of 6-OHDA (0.6, 0.8 μg) failed to yield detectable changes in song (data not shown) even though reduced TH levels were observed in immunoblots with as little as 0.6 μg (Fig. A2-2B).

To prevent oxidation, 6-OHDA was prepared within 30 min of use and kept covered on ice to minimize light exposure. Area X was targeted from the bifurcation of the mid-sagittal sinus, in mm: 5.15 rostral, 1.5–1.6 lateral, and a depth of 3.0–3.3. Injections were delivered every 15 sec.

The total volume was 250 nL for 6-OHDA (n = 9) and 250 or 500 nL for vehicle (n = 11). No effect of vehicle volume on song features was observed, so vehicle-injected birds were pooled. After 5 min, the pipette was slowly retracted and the tip visually inspected for clogging. Following post-operative monitoring, birds were returned to their chambers and recorded until death.

Tissue preparation and immunoblotting

Bilateral micropunches of Area X and outlying VSP and nidopallium (N; Fig. 1C) were obtained, processed and immunoblotted according to Miller et al. [172] but with a PVDF membrane. Post hoc thionin staining of punched sections enabled verification of their anatomical precision (Figure A2-1C). DA biomarkers (Figures A2-2, A2-3) were detected with overnight incubation at 4°C with primary antibodies against TH (Millipore #AB152, rabbit 1:500, 1:1500 and DARPP-32 Abcam #ab40801, rabbit 1:10,000, 1:30,000 dilution;[149]). A primary antibody to GAPDH (Millipore #MAB374, mouse 1:10,000) served as a loading control because neither the Area X protein nor mRNA levels are affected by this behavioral protocol [33,51]. Following TBST washes, blots were probed with HRP secondary antibodies: anti-rabbit IgG (1:2000 – TH, 1:10,000 – DARPP-32) and anti-mouse IgG (1:6000–1:10,000 – GAPDH; Amersham Pharmacia Biotech) for 2 h at room temperature then washed. Blots were developed using chemiluminescence and imaged (Typhoon scanner, or Bio-Rad system) with quantification done in Quantity One (Bio-Rad) by an experimenter blind to the behavioral condition. Densitometric analysis of bands on the immunoblots was as previously described [33,51]. Briefly, a rectangular band was drawn to encapsulate the signal of interest deemed a "raw" value (the "volumetric" measurement in Quantity One) and a same-size rectangular band was placed in the lane above or below the band to subtract

the “background” signal. This yielded a corrected value. Corrected values were obtained for TH, DARPP-32 and GAPDH. Corrected values for TH and DARPP-32 were then divided by a corrected GAPDH value per lane to control for equal protein loading. Protein values reported in Figures A2-2, A2-3 represent these normalized values. Results were independently confirmed, using NIH Image J and the same procedure above was based upon densitometric measurements of the bands.

Immunohistochemistry

Three adult male zebra finches were injected with 1.2 μg of 6-OHDA in Area X of one hemisphere and with vehicle in the other. Brains were collected following a transcardial perfusion of warmed saline followed by 4% room temperature paraformaldehyde in Phosphate Buffer Saline (PBS) on the morning of day 5 (0HR NS). Fixed brains were cryoprotected in 20% sucrose overnight, then frozen in dry ice and sectioned at 30 μm on a cryostat (Microm). The targeting of Area X was visually verified while sectioning by identification of the electrode track. The injection of 6-OHDA results in a brown discoloration in Area X that is visible to the naked eye. Within a given bird, TH immunostaining was compared between the vehicle-injected versus 6-OHDA injected side. The tissue was double-labeled with TH and the neuronal marker NeuN to confirm that Area X neurons were preserved despite poisoning TH nerve terminals.

Tissue sections were processed as follows: Hydrophobic borders were drawn on the slides, using a pap pen (ImmEdge, Vector Labs) followed by 3 × 5 min washes in TBS with 0.3% Triton X (Tx). To block non-specific antibody binding, the tissue was then incubated for 1 h at room

temperature with 5% goat serum in TBS/0.3% Tx then 3 × 5 min washes in 1% goat serum in TBS/0.3% Tx were performed. Primary antibodies to TH (Millipore rabbit 1:500), NeuN (Millipore #MAB377, mouse 1:500) were incubated in a solution of 1% goat serum in TBS/0.3% Tx overnight at 4°C. A "no primary antibody" control was included. The next day, 5 × 5 min washes in TBS/0.3% Tx were performed and sections were incubated for 4 h at room temperature in fluorescently labeled secondary antibodies (Molecular Probes/Life Technologies, 1:1000, goat anti-rabbit 488 #A11034; goat anti-mouse 546 #A11031). Following incubation, 5 × 5 min washes were performed in TBS with filtered TBS used in the last two washes. Slides were then coverslipped in ProLong Anti-Fade Gold mounting medium (Molecular Probes, #P36930), viewed on a confocal microscope (Zeiss LSM 510) using Zeiss LSM software and analyzed, using Adobe Photoshop. In Adobe Photoshop, mean intensity values for TH fiber staining were obtained by measuring the same size rectangular area within Area X of both hemispheres. Mean values for 6-OHDA were then divided by vehicle values to obtain a percentage of TH fiber loss.

Results

Validation of DA biomarkers

A polyclonal antibody made against TH (498 aa; GenBank: AAA42258.1) from rat pheochromocytoma was used for detection. This antibody detects TH depletion in rat basal ganglia following 6-OHDA injection into the medial forebrain bundle [196]. The immunizing peptide shares 76% identity to the predicted 491 amino acids in zebra finch TH (~55 kD; GenBank: XP_002198967). In immunoblots, a robust signal was observed at similar molecular weights across multiple basal ganglia subregions in finch and mouse tissues (Fig. A2-2A). Signals were substantially reduced in the finch nidopallium, consistent with the reduced dopaminergic innervation to this area relative to the basal ganglia in intact birds [197]. A polyclonal antibody against DARPP-32, previously used in zebra finches [149], detects protein signal at the expected molecular weight in region-specific areas of both species (~32 kD; Figure A2-2A).

Immunoblots revealed that bilateral injection of either 0.6 μg or 1.2 μg of 6-OHDA into Area X reduced TH levels, with a more pronounced effect at the 1.2 μg dose (Fig. A2-2B). Additionally, fluorescent immunohistochemistry was conducted on fixed coronal tissue sections from birds receiving a unilateral dose of 6-OHDA injected in Area X of one hemisphere and vehicle in the other. In a representative section, intact, densely packed TH positive fibers were detected throughout Area X in the vehicle-injected hemisphere compared with decreased TH fiber staining (by ~30%) in Area X in the 6-OHDA injected hemisphere (Figure A2-2C). NeuN staining

confirmed that only afferent fibers were lost as the neuronal cell bodies were still present in the 6-OHDA injected Area X (Figure A2-2C).

6-OHDA administration into Area X reduces levels of TH but not DARPP-32 protein

A bilateral injection of 1.2 μg of 6-OHDA into Area X significantly reduced TH signal relative to signals in vehicle-injected birds (Figures A2-3A and A2-3B; n = 4/group; mean ± SE: vehicle 1.47 ± 0.19 vs. 6-OHDA 0.61 ± 0.21; resampling mean difference P = 0.006). In the outlying VSP from these same birds, no such reduction was observed (Figures A2-3C and A2-3D, mean ± SE: vehicle 1.63 ± 0.22 vs. 6-OHDA 1.90 ± 0.14; P = 0.22) indicating that the neurotoxin was confined to Area X. DARPP-32 levels were unaffected in either region (Figure A2-3, Area X: P = 0.56; VSP: P = 0.87), suggesting that the 1.2 μg dose damages presynaptic terminals without affecting at least one postsynaptic marker nor inducing MSN cell death, consistent with the NeuN staining described above.

6-OHDA administration into Area X decreases vocal variability during UD but not FD song

UD song was compared pre and post-bilateral injection of a 1.2 μg dose of 6-OHDA (n = 7 birds) or vehicle (n = 11 birds) into Area X. Compared to vehicle controls, the 6-OHDA injected birds displayed decreased vocal variability in several acoustic features within the bird's song, reflected by increased mean accuracy scores for individual syllables when comparing effect sizes (Figure A2-4A, resampling independent mean differences, P = 0.0298). Syllable exemplars also illustrate the increased accuracy (i.e., stereotypy) post-6-OHDA injection (Figure A2-4B).

Reduced variability (CV) in mean frequency was noted for the effect size plots (Figure A2-4C, resampling independent mean differences, $P = 0.019$). Trends for reduced variability in syllable duration, frequency modulation (FM), entropy, and pitch goodness were also observed in the effect size analysis and met significance for these first three measures in the raw data.

In these same birds, using the effect size comparison, no significant effects of 6-OHDA were detected on mean and CV scores for syllable features in FD song (Figures A2-5A and A2-5B). An evaluation of the raw scores for pre versus post-6-OHDA injection during FD song, revealed significance for mean pitch goodness that was not present in the vehicle-injected birds.

No effects of 6-OHDA on social-context-dependent song differences

Acoustic features in zebra finch song are differentially modulated depending on social context: Syllable subtypes known as harmonic stacks have higher variability in fundamental frequency (FF) reflected as higher CV scores in UD versus FD song [6]. Comparison of the FF CVs for 28 syllables in our own data prior to vehicle or 6-OHDA injection is consistent with prior reports. These syllable subtypes which are modulated by endogenous DA [151,195], were not altered here by 6-OHDA; social-context-dependent differences persisted post-injection. Effect size plots also indicate no significant effect of 6-OHDA on these CV scores (Figure A2-6A) and the UD versus FD differences in FF variability were preserved (Figure A2-6B). However, a power analysis of these syllable subtypes revealed that the ability to detect differences due to the drug is only 10%.

An investigation of all syllable subtypes (harmonics included) revealed that pre-surgery, FD song had higher syllable self-similarity and accuracy scores compared to UD song (Figure A2-7A, "pre"). Pitch and entropy changes were more variable during UD versus FD song in the pre-surgery group (Figure A2-7C). Following 6-OHDA or vehicle injection, these differences were preserved. Overall, the mean and CV scores for song features indicated no attenuation of social-context-dependent differences with 6-OHDA (Figures A2-7B and A2-7C). Unexpectedly, one feature, mean pitch goodness (Figure A2-7B), for the vehicle group was greater in UD than FD.

6-OHDA administration and non-song motor features

An observer blind to the treatment scored the frequency of non-song behaviors (eating, drinking, alarm calls, grooming, flying, beak-wiping, preening and following the female) in a subset of the vehicle ($n = 4$) and 6-OHDA injected birds pre versus post-injection for both UD and FD states ($n = 6$). Comparing the pre versus post-surgery vehicle group, no changes in behavior were detected. Following 6-OHDA injection, the only feature that changed was increased beak-wiping during FD (resampling paired difference, $P < 0.05$), but this behavior was rare, limited to 1–2 wipes/session.

Discussion

Bilateral injection of a 1.2 μg dose of 6-OHDA targeted to Area X reduced DA signal within this song control nucleus, as measured by quantification of TH levels on immunoblots. No changes in postsynaptic DARPP-32 levels within Area X were detected. The accuracy of our targeting was validated by the lack of change in TH and DARPP-32 signals in outlying VSP. A relatively novel aspect of our approach was to obtain tissue punches from both rodent and finch basal ganglia regions in order to quantitatively measure DA signals from entire nuclei via immunoblotting. This method can provide a more complete picture of overall DA loss allowing for multiple animals to be analyzed on one blot. A qualitative immunohistochemical image of TH fiber loss in Area X due to 6-OHDA injection (Figure A2-2C) supports the more quantitative results obtained from the immunoblots. Based on the unique aggregation of birdsong control neurons within their surrounding brain regions, this approach further offers the opportunity to manipulate DA levels specifically within a vocal control region of the basal ganglia.

Basal ganglia tissue and song measurements were sampled over an acute phase of treatment, 4–5 days following injection of 6-OHDA, in order to quantify the TH loss and subtle changes within the bird's song over an early time window. This period was selected to model the effects of early DA loss on vocal symptoms due to the loss of TH-positive axons in the striatum prior to death of midbrain DA cells [198,199]. In rodent models of early disease, synaptic degeneration and loss of TH-positive axons in the striatum can be detected as early as 24 h post-injection and becomes more marked by 5 days. This loss of TH-positive axons precedes the

retrograde degeneration and death of the midbrain DA cells that takes place over weeks to months[199].

The main differences following 6-OHDA injection in rat versus finch lie in the severity of the lesion and associated vocal symptoms. In the rodent literature, the striatal TH loss varies in severity depending upon the site of injection (lateral ventricles, median forebrain bundle) and the 6-OHDA dosage used, with higher doses resulting in a more complete loss of TH fibers [194,199,200]. The associated degradation in the vocal signal affects some, but not all, features such as frequency-modulated complex calls [201,202]. In contrast, here in zebra finch Area X, reduction in the DA biomarker TH results in subtle loss of UD song variability likely due to the lower 1.2 μg dose of 6-OHDA. The higher 7 μg dose used in rodents [194] proved lethal to finches, and doses >1.2 μg can induce a lesion in Area X (data not shown).

We predicted that UD song would be more sensitive to 6-OHDA effects than FD song, given that DA levels in Area X are already quite low during UD singing [191]; a small loss of DA would thus proportionally affect more of the UD than the FD-associated levels. Confirming this prediction, decreases in variability were detected in syllable accuracy scores (syllables became more similar across renditions) and reduced variability in mean frequency. Strong trends for decreased variability in individual features that comprise the accuracy score calculation were detected post-6-OHDA injection for syllable duration, frequency modulation, entropy, and pitch goodness that were not detected in vehicle-injected birds. These trends were evident for both the effect size and the raw data.

The effect sizes (Figures A2-4, A2-5, A2-6, A2-7; see Methods) of each treatment or social context on every syllable, provide a direct statistical comparison between conditions. In addition to effect size, we also examined the results of statistical tests where song data were paired, and analyses were performed within a condition. The two methods show overall agreement; in a few instances significance is observed when the data are viewed as pairs but not when calculating the effect size. While both results are valid, we focus on the effect size because this transformation enables the direct comparison between two conditions (e.g., vehicle vs. 6-OHDA); a comparison not possible when data are analyzed as pairs (e.g., pre-vehicle vs. post-vehicle). Additionally, calculating differences between raw values when data are paired allows large and small values to heavily skew the overall difference between conditions, which is alleviated by the normalization in calculating effect size. Plotting the effect size also clearly indicates the magnitude of the change resulting from the experimental manipulation, or condition.

Using these analyses, 6-OHDA-induced changes were detected in UD but not FD song. The reduction in UD song variability following DA-depletion is reminiscent of vocal changes in human PD in which reduced vocal variability is evident in a wide range of symptoms including breathy, soft, rough and monotonous voice, impairment in coordination of orofacial articulators, and fluency (reviewed in Sapir [203]). Symptoms such as monotonous voice occur early in the disease [204] before the large-scale loss of DA cells in the SNc. In contrast, because FD song is associated with elevated Area X DA levels [191], the DA loss associated with the 1.2 μg dose may be proportionally too low to impact FD. Interestingly, the lack of detectable changes during FD song is reminiscent of the observation that PD patients perform better when externally cued, for

example, by a speech-language pathologist [205]. The apparent lack of 6-OHDA's effect on FD may be a consequence of low power (see below) and/or reflect compensatory mechanisms related to external cues that override neuropathological deficits.

We next evaluated the effect of 6-OHDA induced reduction in DA signals in Area X on social context-dependent differences normally evident during UD versus FD song. Zebra finch FD song is characterized as being more stereotyped, based upon the analysis of one particular syllable type [6]. Specifically, syllables with low frequency-modulation, known as harmonic stacks, have reduced variability in FF from rendition to rendition in FD song compared to UD [6]. Pharmacological blockade using a D1 receptor antagonist in Area X causes these harmonic syllables to become more variable during FD song [151]. The higher variability in FF in UD versus FD song was observed here prior to 6-OHDA injection, replicating prior findings. Unexpectedly, there was no detectable effect of mild 6-OHDA-mediated DA depletion on variability scores. Power analysis indicated that the ability to detect any such effect was limited by the low number of harmonic stacks available to analyze. These make up only a subset of all syllable types unless birds are selectively bred to obtain multiple harmonic syllables in their motifs.

To date, one other study has examined natural differences in UD versus FD song at the syllable level, reporting that subsyllabic elements in the bird's song are more spectrally similar in FD than UD [151]. Our acoustic analysis examined a wider range of song features between the two social contexts and found that pre-surgery, FD song has higher self-similarity and accuracy scores across multiple renditions, supporting the previous literature that FD song is more stereotyped compared to UD. Although 6-OHDA affected small changes in UD song features

observed pre versus post-surgery (Figures A2-4A and A2-4C) it was not sufficient to cause UD song to fully resemble FD song.

Receptor-mediated mechanisms may underlie 6-OHDA effects on UD song. In rats, D1 and D2 receptor activation modulate their ultrasonic vocalizations both separately and synergistically [196]. The function of these receptors in vocalizations may be differentially altered with 6-OHDA administration. For example, injection into the SNc of rodents results in elevated striatal D2 receptor mRNA levels, whereas D1 receptor mRNA is reduced [206]. In adult male zebra finches, D2 receptors appear to be the prominent basal ganglia subtype [207]. Yet, as mentioned, a D1 receptor antagonist causes FD song to resemble UD song [151]. Because many Area X MSNs co-express both D1 and D2 receptors [207], discerning the subtype specific effects of 6-OHDA injection will be challenging. Moreover, the traditional view that activation of D1 receptors promotes excitability in the direct pathway whereas D2 suppresses it in the indirect pathway has been revised to recognize that the mammalian striatopallidal pathways and projections are anatomically and physiologically intertwined [208,209].

Although it is likely that DA receptors are abnormally activated following 6-OHDA administration, altered adrenergic signaling cannot be excluded. DA can bind to alpha-2 type adrenergic receptors, which are abundant in Area X [210]. Ongoing work is aimed at assessing any alterations in these receptor levels following 6-OHDA administration. Norepinephrine (NE) from the locus coeruleus can also bind to adrenergic receptors, but there is sparse NE in Area X based on HPLC measurements [197] and immunostains for dopamine beta hydroxylase, the biosynthetic enzyme for NE [211].

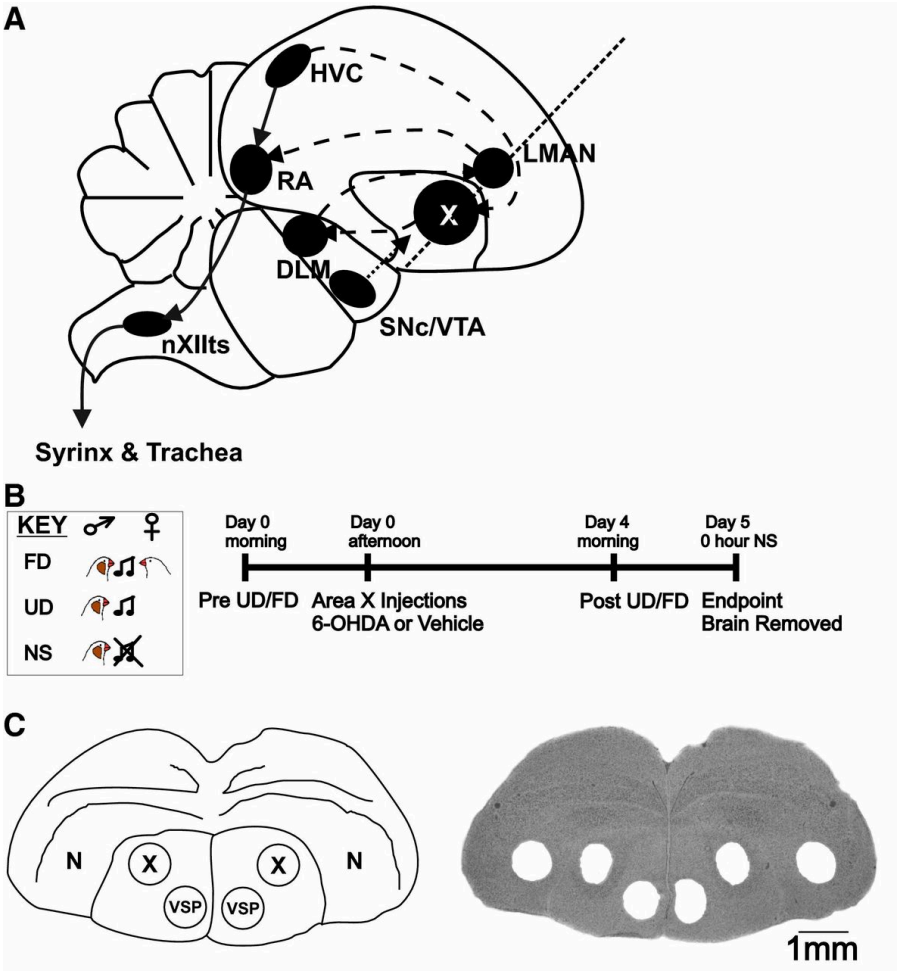
The progressive neuropathology associated with PD in the brainstem and cortico-basal ganglia circuits [212] contributes to voice and speech symptoms, but the underlying neural mechanisms are not well-understood. The 6-OHDA zebra finch model described here provides a convenient entry point to examine the impact of mild DA loss on synaptic mechanisms and song. By studying song circuitry, inferences can be made regarding neural mechanisms underlying early vocal changes in human PD. Indeed, the special song-dedicated nucleus, Area X, shares more similar gene expression patterns with the putamen, a speech active region in humans than with these areas in non-vocal learning birds and primates [10]. Future investigations will use 6-OHDA as a tool to identify nigrostriatal genes sensitive to DA depletion [51,213]. Combining genetic, physiological, and behavioral approaches in the well-characterized vocal circuitry of the songbird will advance understanding of circuits that drive their vocal apparatus with implications for humans.

Acknowledgements

We thank Debora Lee and Jingwen Yao (UCLA) for behavioral analyses, Jonathan B. Heston (UCLA) for input on the figures; Stephanie Munger, Joshua L. Ritter, Patty Jansma, and Professor Nicholas Strausfeld (U. Arizona) for technical contributions.

Figures

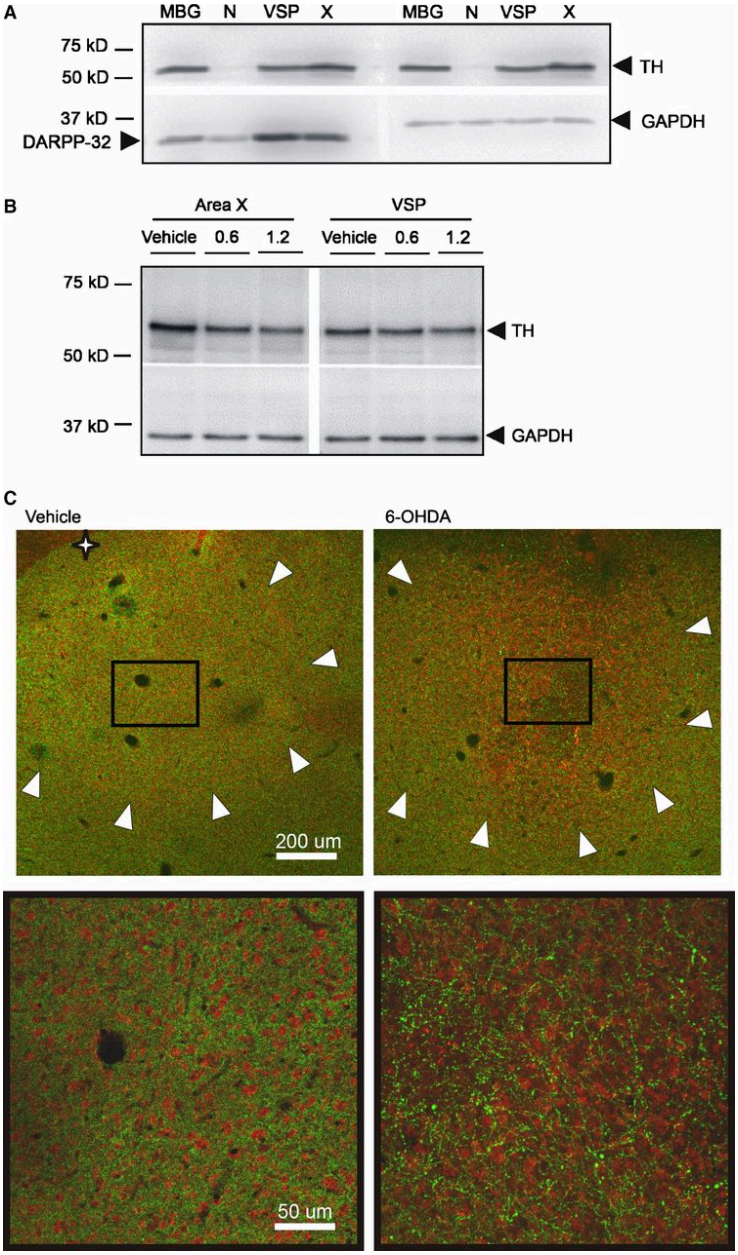
Figure A2-1: Neuroanatomy of the song circuitry and experimental timeline.



a) The song control system mainly consists of two interconnected loops: the vocal production pathway (solid lines) containing cortical nuclei HVC (proper name) and the robust nucleus of the arcopallium (RA); the anterior forebrain pathway (dashed lines) including basal ganglia Area X (X), the dorsolateral division of the medial thalamus (DLM) and the cortical lateral magnocellular nucleus of the anterior nidopallium (LMAN). Area X receives DA input (dotted arrow) from the substantia nigra pars compacta (SNc) and ventral tegmental area (VTA) [168]. A dotted line

indicates the coronal plane of section shown for (C). Modified from [172]. nXIIts – tracheosyringeal portion of the hypoglossal motor nucleus. Other abbreviations in text. **b)** Experimental timeline. The key (Fig. 1B), represents the behavioral contexts for 2 h of undirected (UD) song, female directed (FD) song and 0 h non-singing (NS), the experimental endpoint. In the case of insufficient singing, this timeline was adjusted ±1 day for pre-surgery and post-surgery song collection. **C)** Schematic of male zebra finch coronal brain section (left) indicates anatomical regions and micropunches in the thionin-stained section (right).

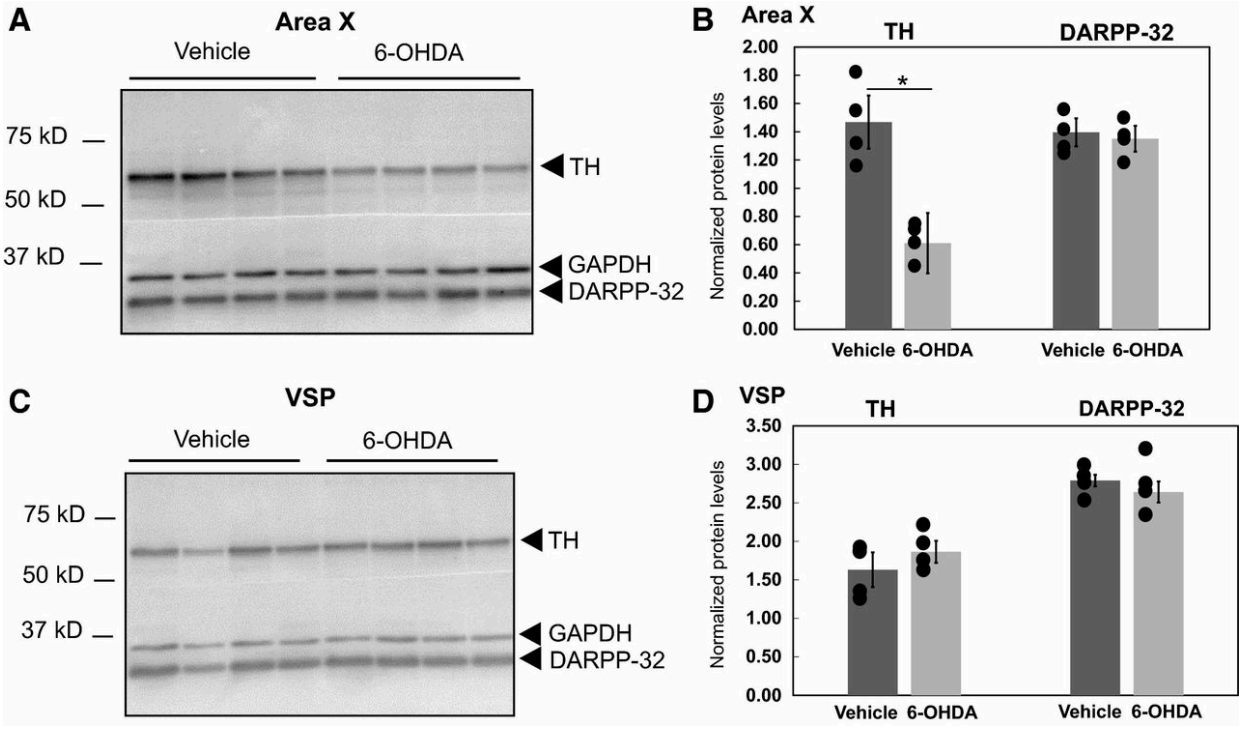
Figure A2-2: Tissue measurements of DA biomarkers.



a) Immunoblot (40 μg protein/lane) from Area X, VSP, nidopallium (N), and mouse basal ganglia (MBG) lysates. Signals are at the expected molecular weights (kD) and show the expected reduction in TH signal within nidopallium. **b)** Immunoblot (15 μg protein/lane) from Area X and VSP lysates. Compared to vehicle (lane 1: 2.82, normalized protein levels), TH signal appeared

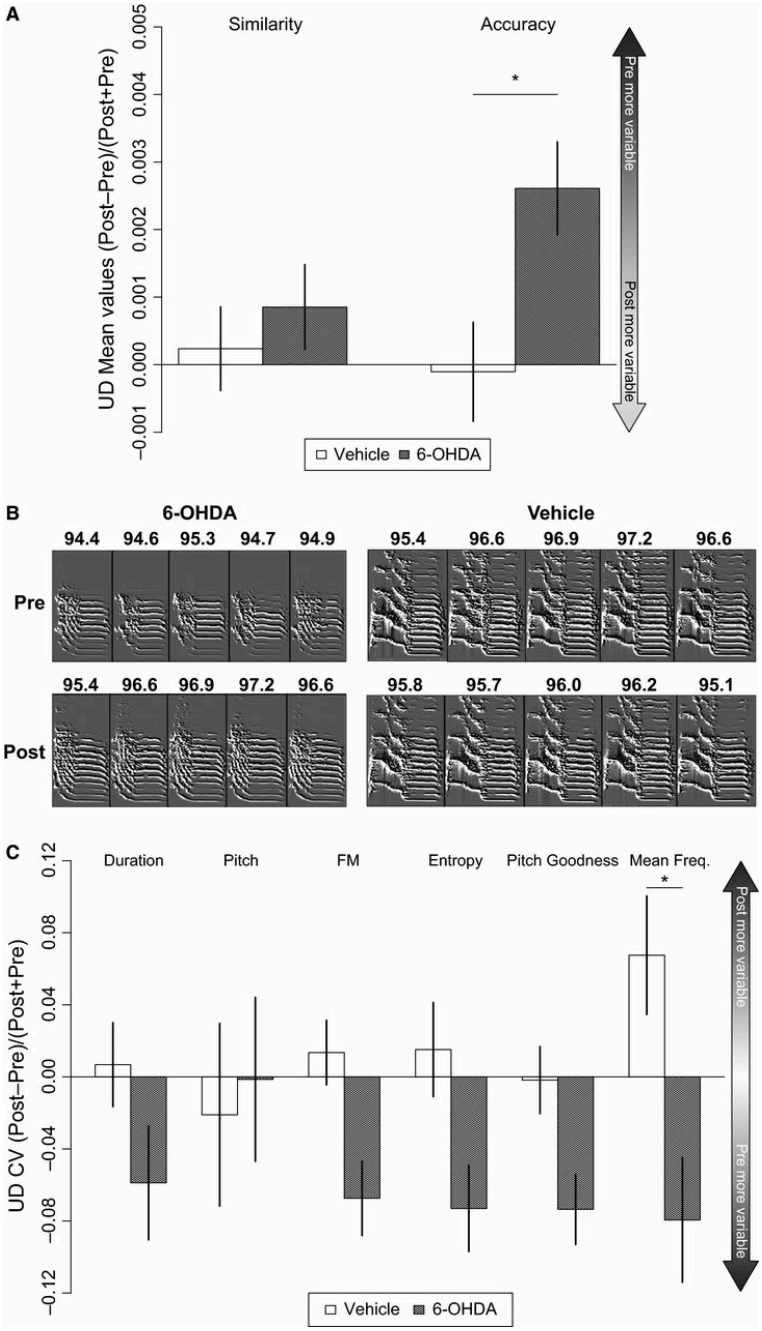
reduced in Area X following both 0.6 μg (lane 2: 1.75) and 1.2 μg (lane 3: 1.03) doses of 6-OHDA, with more substantial reduction at the higher dose. TH signals in VSP exhibited less change as expected given the targeted injection to Area X (lanes 4–6; vehicle: 1.31; 0.6 μg: 1.06, 1.2 μg: 0.93). **c)** Decreased TH immunostaining in Area X following 6-OHDA injection. Photomicrographs show double-labeling for TH positive fibers in green and NeuN, a neuronal marker, in red. The star indicates the striato-pallidal border – beyond this border, the nidopallium lacks the density of TH fibers. Arrowheads outline Area X (top); rectangle highlights inset shown below at higher magnification. There are fewer TH fibers (green) in the 6-OHDA injected Area X compared to vehicle-injected but the density of NeuN staining (red) indicates that Area X neurons are preserved.

Figure A2-3: 6-OHDA reduces TH signal in Area X but not VSP.



a–b) Immunoblot (A, left; 15 μg/lane) shows decreased TH signal in 6-OHDA- versus vehicle-injected birds, as quantified in the accompanying graph (B, right). No change in DARPP-32 signal was detected. Quantification shows means (bars), standard error (plungers), and individual bird values (circles). **c–d)** Immunoblot (C, left; 15 μg/lane) with accompanying graph (d, right) indicates no change in TH and DARPP-32 signals in the VSP for the same birds as in a–b.

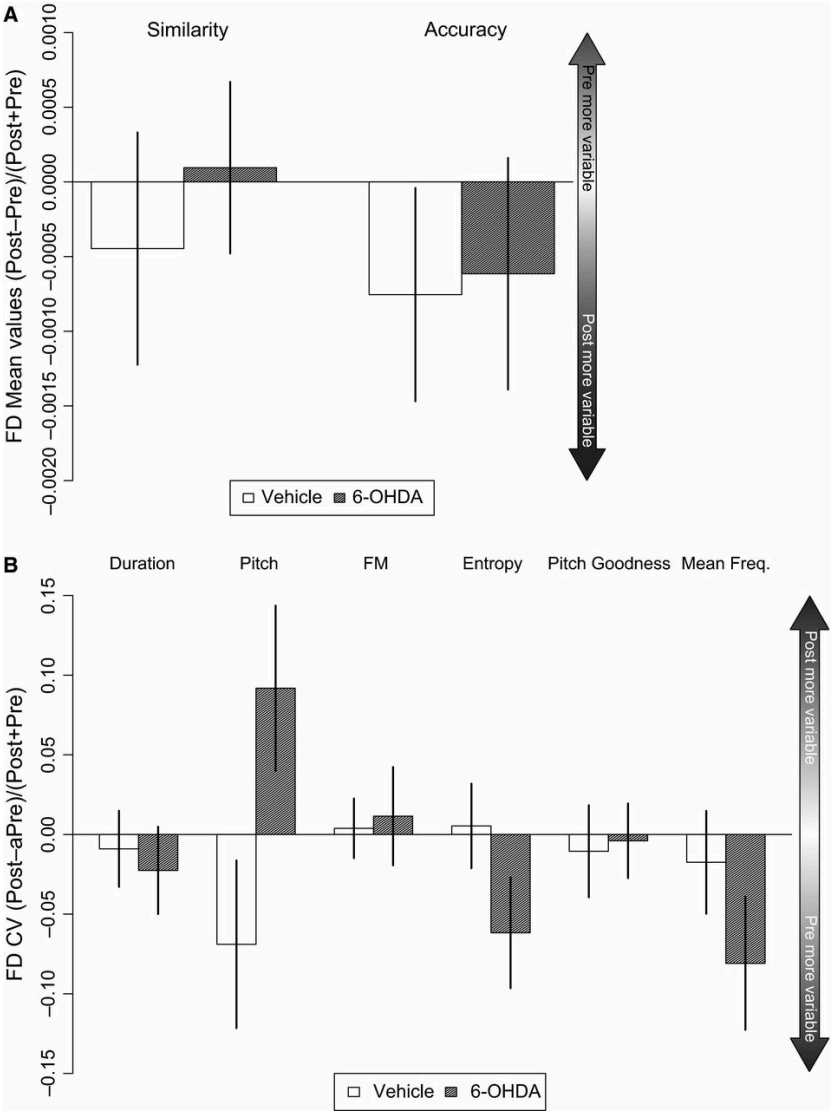
Figure A2-4: UD features become less variable with 6-OHDA injection in Area X.



Effect sizes for the 6-OHDA versus the vehicle condition display the median score with standard error bars from all syllables for each bird. **a)** Mean syllable self-accuracy, denoted by the higher

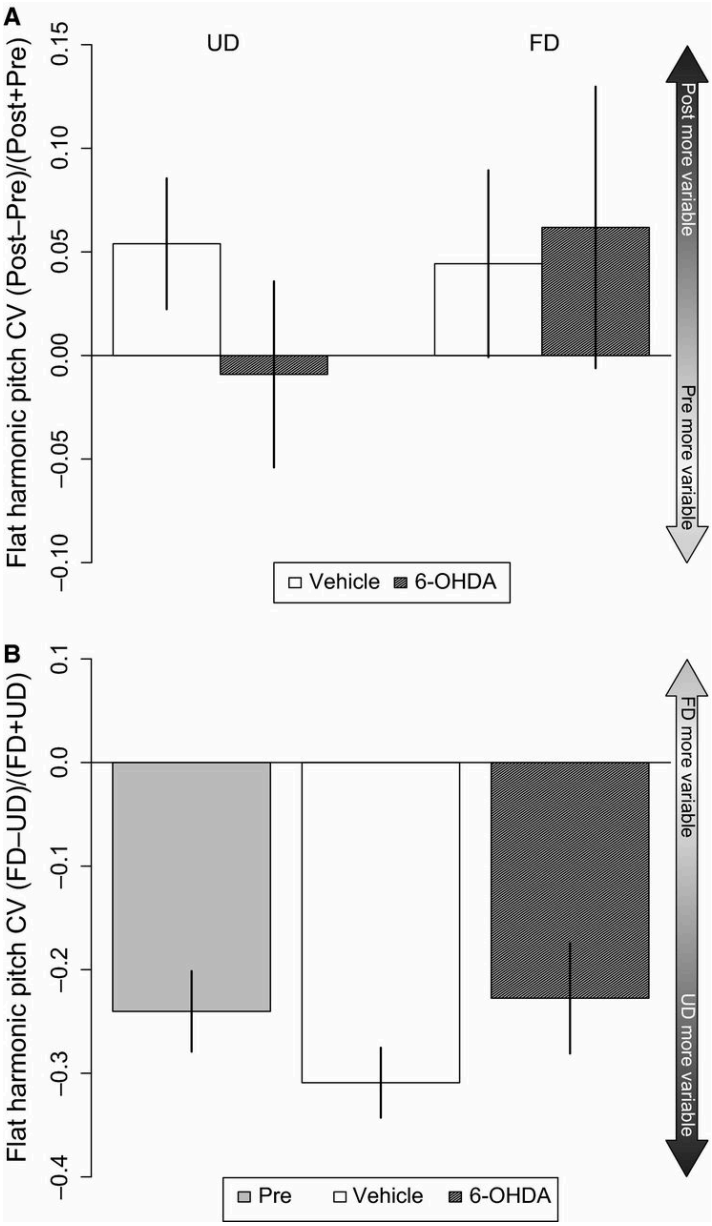
bar, is significantly greater post-6-OHDA injection (*) compared to vehicle-injected birds. Self-similarity is not affected by treatment. **b)** Consecutive renditions of the same syllable pre versus post-6-OHDA injection or vehicle with mean accuracy scores. Following 6-OHDA injection, the syllable increases in self-accuracy over multiple renditions, not observed in vehicle-injected birds. **c)** Duration, frequency modulation (FM), entropy, and pitch goodness show a trend for decreased variability in the post-6-OHDA injected birds compared to vehicle-injected controls. Variability in mean frequency (coefficient of variation, CV) is significantly less (*) following injection of 6-OHDA versus vehicle.

Figure A2-5: FD features are not affected by 6-OHDA injection in Area X.



Effect sizes for each treatment represent the median score with standard error bars from all syllables for each bird in vehicle or 6-OHDA-injected groups. Mean **(a)** and CV scores **(b)** in FD song were not significantly affected by 6-OHDA injection.

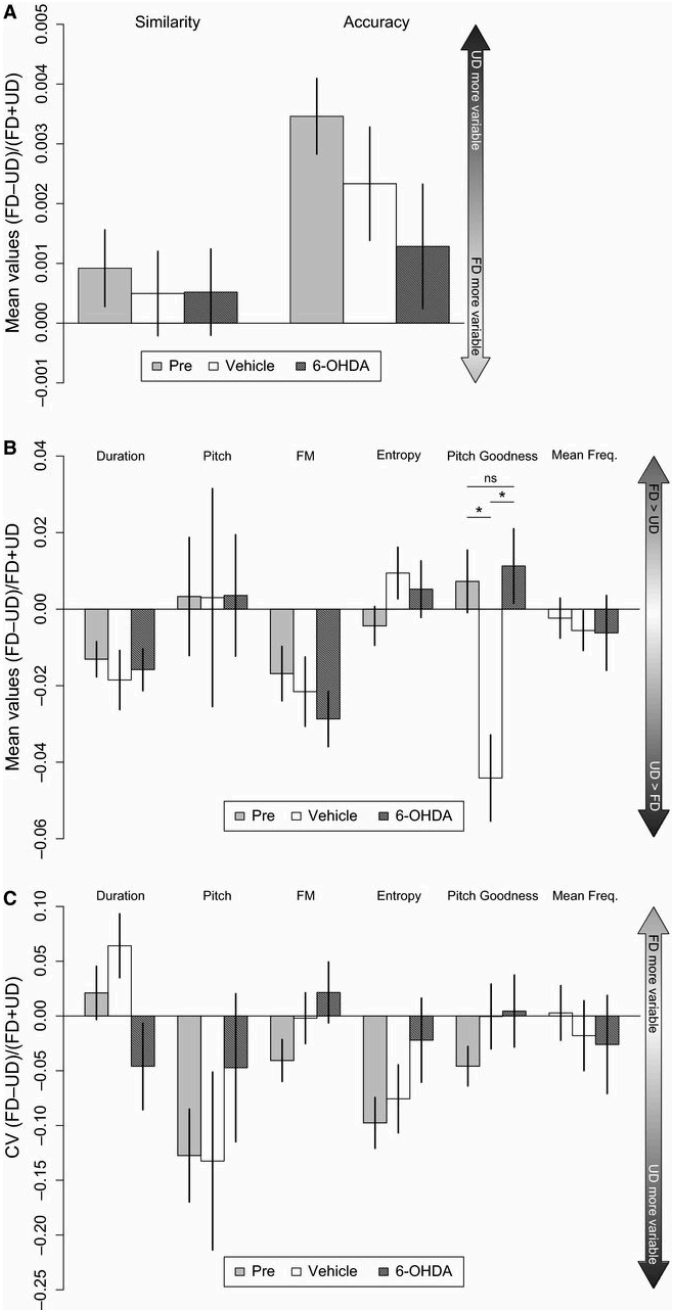
Figure A2-6: Social-context-dependent differences for harmonic syllables are not affected by 6-OHDA injection in Area X.



A comparison of UD versus FD song features following vehicle or 6-OHDA injection was made using the effect size plots with some refinement: The "Pre" bar represents all pre-surgery syllables combined from both pre-treatment groups. The "vehicle" bar represents comparisons between

post-vehicle FD and post-vehicle UD whereas the "6-OHDA" bars are comparisons between post-6-OHDA FD versus post-6-OHDA UD song. **a)** No significant difference was observed between vehicle and 6-OHDA on flat harmonic syllable CV in UD (left) or FD (right) song. **b)** Before surgery ("Pre") UD song has greater pitch variability than FD song. Following surgery ("6-OHDA" and "Vehicle"), pitch variability continues to be greater in UD song.

Figure A2-7: Social-context-dependent mean and CV scores for all syllable types are not affected by 6-OHDA injection in Area X.



a) Syllable self-similarity and self-accuracy are greater in FD than UD song before surgery ("Pre"; positive bars). Greater self-similarity and self-accuracy indicate less variability across multiple

renditions of the same syllable. Neither injection with vehicle nor 6-OHDA caused a significant change in this trend. **b)** The effect of injection with vehicle or 6-OHDA did not cause a significant change in mean scores between FD and UD songs except for pitch goodness in which UD became greater than FD in vehicle birds (*). **c)** The effect of injection with vehicle or 6-OHDA did not cause a significant change in the modulation of variability (as measured by CV) between FD and UD songs. Positive bars indicate greater variability in FD while negative bars indicate greater variability in UD. In the pre-surgery group, the negative bars of greater magnitude for pitch and entropy indicate that these features were more variable for UD versus FD song.

Appendix 3: Mice with Dab1 or Vldlr Insufficiency Exhibit Abnormal Neonatal Vocalization Patterns

Elizabeth R. Fraley, Zachary D. Burkett, Nancy F. Day, Benjamin A. Schwartz, Patricia E. Phelps, & Stephanie A. White

Statement of Contribution

The Reelin signaling pathway is associated with autism spectrum disorder and a number of its members are regulated by vocal behavior in the zebra finch. This study is an examination of the reelin receptor Vldlr, an intracellular signaling molecule Dab1, and their contributions to the innate vocalizations of mice. Like the work described in Appendix 3, I contributed mainly to the analysis of vocal behavior data. My work is reflected in figures A3-2, A3-3, A3-5, A3-6, A4-7, and A3-8 where we describe how deletion of these genes affect the distribution and sequencing of calls.

Abstract

Genetic and epigenetic changes in components of the Reelin-signaling pathway (RELN, DAB1) are associated with autism spectrum disorder (ASD) risk. Social communication deficits are a key component of the ASD diagnostic criteria, but the underlying neurogenetic mechanisms remain unknown. Reln insufficient mice exhibit ASD-like behavioral phenotypes including altered neonatal vocalization patterns. Reelin affects multiple pathways including through the receptors, Very low-density lipoprotein receptor (Vldlr), Apolipoprotein receptor 2 (Apoer2), and intracellular signaling molecule Disabled-1 (Dab1). As Vldlr was previously implicated in avian vocalization, here we investigate vocalizations of neonatal mice with a reduction or absence of these components of the Reelin-signaling pathway. Mice with low or no Dab1 expression exhibited reduced calling rates, altered call-type usage, and differential vocal development trajectories. Mice lacking Vldlr expression also had altered call repertoires, and this effect was exacerbated by deficiency in Apoer2. Together with previous findings, these observations 1) solidify a role for Reelin in vocal communication of multiple species, 2) point to the canonical Reelin-signaling pathway as critical for development of normal neonatal calling patterns in mice, and 3) suggest that mutants in this pathway could be used as murine models for Reelin-associated vocal deficits in humans.

Introduction

Reelin is a large secreted glycoprotein that has numerous nervous system functions including regulating neuronal migration, neuronal excitability, and dendritic morphology[214-219]. Murine *reeler* mutants (*Reln*−/−) do not express Reelin and exhibit a characteristic phenotype of a reeling gait, disorganization of laminated structures including the neocortex, cerebellum, and hippocampus, and a reduction in cerebellar volume[214,220-223]. When Reelin binds to Very low-density lipoprotein receptor (Vldlr) and/or Apolipoprotein receptor 2 (Apoer2), this initiates binding of Disabled-1 (Dab1) to the internal domain of the receptors[224,225]. Dab1 is then phosphorylated at critical tyrosine residues by Src-family kinases to influence a wide array of downstream effectors[219,226,227].

RELN has been identified as a risk allele for autism spectrum disorder (ASD) in multiple populations [228-236]. Polymorphisms throughout RELN include variants in both coding and non-coding regions. Changes leading to an expansion in GGC repeats in the 5′ region reduce RELN expression levels and confer ASD risk in some cases24. Reelin protein (RELN) is low in post-mortem brain tissue of ASD patients compared to controls [237]. Additionally, RELN mRNA is low in the cerebellum and cortex of these patients [237]. Epigenetic down regulation of RELN via increased methylation of its promoter is also linked to increased ASD risk [238]. Intriguingly, DAB1 polymorphisms are associated with ASD risk in the Chinese-Han population whereas RELN polymorphisms are not [239]. These results indicate that not only RELN, but the function of the downstream Reelin-signaling pathway could be involved in the etiology of ASD.

Social communication deficits are a key diagnostic feature of ASD. Autistic symptoms are generally undetected at birth, but instead appear over time and reflect differential developmental trajectories [240]. High risk infants, i.e. children with one or more ASD siblings, and infants later diagnosed with ASD, have altered acoustic features of their cries [241]. At 6 months, their cries are more disordered and of higher pitch compared with those of typically developing children. High risk infants also exhibit abnormal pre-linguistic vocal behavior such as making fewer speech-like vocalizations and more non-speech vocalizations as well as producing fewer consonant types than typically developing peers [242]. Given these observations, examination of the amount and acoustic parameters of infant cries could serve as a tool for early ASD detection.

Genetic causes are linked to 10–25% of ASD cases [243]. Investigation of how these gene mutations alter behavioral phenotypes, and the underlying brain organization and function, are enabled by mouse models [126]. Neonatal mouse pups typically emit ultrasonic vocalizations (USVs) when isolated from the dam which act as a signal for the dam to retrieve and care for the pups [244]. Pups are entirely reliant on the dam during this time (P0-P14), and thus appropriate communication cues are critical to their survival. Isolation USVs first occur at ~postnatal day 4 (P4) and peak around P6-7, before gradually declining at P14 when the pup is able to self-retrieve [245]. Dams preferentially retrieve pups that call more and prefer more elaborate call types [246,247]. Reductions in total calling, delays in the peak calling age, and an altered call repertoire occur in many mouse ASD models [114,132,248-254]. Murine neonatal isolation calls are considered to be more like human baby cries than early speech. Although laboratory mice are not robust vocal learners [107], their vocal behavior can reflect sociability and mechanisms of communication that subserve both learned and unlearned vocalizations.

Because humans with ASD were reported to have low levels of Reelin, the Reln+/− mouse was proposed as a model for ASD [255]. Reln+/− mice have a 50% reduction in Reelin protein and lack the typical neuronal migration deficits seen in Reln−/− mice [256]. Reln+/− mice however, exhibit GAD67 down-regulation in the frontoparietal cortex [257], Purkinje cell loss and hypoplasia of the cerebellum [258] and parvalbumin-positive cell loss in the striatum [259], resulting in changes to cortico-striatal plasticity [258]. These changes are parallel to those in human ASD cases which show the following abnormalities: low GAD67 across brain regions including the frontal cortex [260]; Purkinje cell loss and reduced cerebellar volume [261,262], and altered connectivity and function of the striatum [263,264]. Genetic vulnerability can determine phenotype by interacting with the environment; other studies have examined multivariate conditions (separation, stress, drug/pesticide exposure) that interact with the reduced Reln expression to mirror ASD-like phenotypes [265-269].

To investigate whether or not Reelin deficiency alone creates an ASD-like phenotype, the early vocal behavior of the Reln+/− and Reln−/− mice was characterized [265,270]. Reln+/− mice show a delay in the age of peak isolation USV calling, whereas Reln−/− mice have low calling rates at all measured time points (P2-P12), most likely due to gross motor deficits [265]. Repertoires of P6 pups are altered in a gene-dose dependent manner, with a particularly large expansion of two-syllable call types (see Methods for call type classifications) [270]. Differences in repertoire based on genotype disappear as pups mature (P8-12). These findings indicate a deficit in early vocal communication in Reln+/− mice.

VLDLR is a known target of the language-associated transcription factor FOXP2 in humans [271]. Moreover, vocally regulated gene networks in the zebra finch basal ganglia (Area X) [51] include Vldlr, Dab1, and Reelin; Vldlr is in the same gene module as FoxP2. These observations suggest that the Reelin-signaling pathway is essential for normal vocal development in multiple species.

Here, we examine the early vocal phenotypes of mice with reductions in Vldlr, Apoer2 and Dab1. Findings are then compared with those from Reln+/− and Reln−/− mice [270] in order to attribute changes in vocal development to the canonical Reelin-signaling pathway. Given the greater incidence of ASD in the male population, sex as a contributing factor was also examined.

Methods

Mouse breeding and care

Experiments were approved by UCLA Office of Animal Research Oversight. All animal use was in accordance with the UCLA Institutional Animal Care and Use Committee and complied with American Veterinary Association standards for working with laboratory animals. Mice were maintained on a 12 h light/dark cycle, with *ad libitum* food and water. *Dab1*^{lacZ} mice were a gift from Dr. Brian Howell (Upstate Medical University, SUNY). The mice have a truncation of Dab1 at residue 22 and expression of a fusion of the *lacZ* reporter rendering the protein unable to initiate downstream signaling via phosphorylation at critical residues [69]. These mice were generated as previously described [272] by breeding *Dab1 cKIneo* mice with Meox-Cre germline deleter mice (B6.129S4-*Meox2*$^{tm1(cre)SOR}$/J, Jackson Laboratories, Bar Harbor, ME, USA) and bred into B6;129Sv. The expression of beta-galactosidase is in line with established Dab1 expression in cerebral cortex, cerebellum, and hippocampus [272,273]. *Vldlr*^{tm1Her} mice have a targeted complete deletion of the *Vldlr* gene and were generated in B6;129S7 mice [274]. Double receptor mutants do not have Apoer2 or Vldlr and resemble *Reln*$^{-/-}$ mice [226]. *Dab1, Vldlr,* and *Apoer2* mice were genotyped using PCR as previously described [272,274,275].

Vocal recording

To test isolation calls, mouse pups were removed from the nest, four at a time, and individually placed into sound attenuation chambers for recording. These chambers were constructed from small coolers (Coleman) that were coated inside with soundproof foam

(Soundcoat). An ultrasonic microphone (UltraSoundGate, Avisoft Bioacoustics) was suspended above the pup. Recordings were conducted at P7 and P14 for a total of 15 minutes at each time point. In order to not affect calling patterns of any remaining pups in a given litter, only four (the total number of recording chambers) from each litter were recorded. After recording at P7, pups were tailed for genotyping, and tattooed to enable identification for re-recording at P14. The initial distance between the microphones and the pups was equivalent across chambers at P7 when pups are fairly immobile. At P14, pups are ambulatory so their distance from the microphone varied. Because of this, amplitude measurements were not included in the acoustic analysis. Pups were recorded within the same 2-hour time window each day (light: 14:00–16:00 hr) to avoid circadian effects. Temperature was maintained at 21–22 °C.

Acoustic analysis, quantification and classification

Ultrasonic (20–125 kHz) vocalizations were acquired using a sampling rate of 250 kHz. In order to reduce background noise and focus on ultrasound, sounds with a frequency <40 kHz were high-pass filtered and removed from analysis. Recorded vocalizations were segmented based on amplitude threshold to allow for recording of bouts (Avisoft-SASLab Pro Recorder). Bouts are a series of USVs that occur in rapid succession ($\leq$40 ms between calls) and are surrounded by $\geq$1 second of silence. Recordings were transduced from amplitude traces into spectrograms using Fast Fourier Transform with a transform of 256 points and a time window overlap of 75% (Avisoft Bioacoustics; SASLab Pro). Bouts were then segmented into individual syllables and then processed using VoICE, a semi-automated unbiased clustering mechanism, to classify these calls into categories [69].

Call type categories from the work of Scattoni and colleagues [114] were used and include 9 basic types: 'short' (duration <10 ms), 'downward' (frequency sweeps downward of ≥10 kHz and >10 ms), 'upward' (frequency sweeps upward of ≥10 kHz, >10 ms), 'flat' (<10 kHz of modulation, >10 ms), 'complex' (wave shaped frequency sweep, >10 ms) 'frequency step' (multiple jump containing calls), 'chevron' (inverted U shape frequency with ≥10 kH of modulation), 'harmonic' (multiple jump containing with harmonic stacking), and 'two-syllable' (one-jump containing). Composite call types (those containing no jumps but with harmonic stacking) were collapsed into the harmonic call category; unstructured call types (broadband of ≥40 kHz with no clear single frequency) comprised <1% of the recordings and were not analyzed. Additional call categories of 'doubles', 'triples', and 'miscellaneous' were observed and included. Double and triple calls are comprised of various frequency sweeps that occur in rapid succession, being separated by ≤10 ms. These were rare and considered together as a single call type. Miscellaneous call types did not fit into any of the groups described previously, and may represent emerging novel types. The sequence of the calls, referred to here as 'syntax', was also assessed. Syntax similarity, syntax entropy and repertoire correlation analyses were performed as described previously [69].

Statistical methods

Where possible, resampling statistical tests were used because this methodology makes no assumptions about the data distribution. Call counts were quantified and analyzed using two-way ANOVA followed by individual 2-tailed t-tests with as follows: Once call classifications were determined, we noticed a high degree of variability between individual pups of the same genotype.

To overcome this variability, we normalized each raw call count in each category to the total number of call counts per animal. These normalized values were used to create pie charts. In order to assess statistical differences in repertoire between groups, call count categories of each animal were rank transformed and then resampled 10,000 times to determine the median rank and 95% confidence interval (CI) for each genotype. Only measures with non-over lapping CIs were considered to differ. Syntax similarity scores, syntax entropy scores, and repertoire correlation were also subjected to one-way ANOVA followed by *post hoc* 2-tailed t-tests with Welch's correction.

Results

Amount of calling depends on Dab1 genotype at P7 but not at P14

To test for an early social communication deficit, Dab1 deficient mouse pups were recorded (Figure A3-1A). The number of calls produced by male and female Dab1 pups of each of the 3 genotypes was quantified ($Dab1^{+/+}$ N = 21, $Dab1^{+/lacZ}$ N = 23, $Dab1^{lacZ/lacZ}$ N = 11). A significant effect of genotype on calling behavior was observed at P7, with no effect of sex (Figure A3-1B; two-way ANOVA, sex effect p = 0.308, genotype effect p = 0.005, interaction p = 0.671). As call number did not differ based on sex, data from both sexes were pooled. At P7, $Dab1^{+/+}$ mice called the most and their call counts were significantly greater than those of the $Dab1^{lacZ/lacZ}$ mice (t-test; p = 0.0001). The $Dab1^{+/lacZ}$ mice called more than the $Dab1^{lacZ/lacZ}$ mice (t-test, p = 0.004). The $Dab1^{lacZ/lacZ}$ mutants made the least number of calls at this time point, a result that may reflect their severe motor deficits. Thus, at P7, a Dab1 gene-dose dependent effect on calling amount was evident, with no effect of sex.

At P14, the amount of calling was relatively low and similar across all three genotypes (Figure A3-1C; $Dab1^{+/+}$ N = 17, $Dab1^{+/lacZ}$ N = 19, $Dab1^{lacZ/lacZ}$ N = 9; two-way ANOVA; genotype effect p = 0.545, sex effect p = 0.400, interaction p = 0.296, NS). The amount of calling by $Dab1^{+/+}$ mice did not differ from $Dab1^{+/lacZ}$ mice (t-test, p = 0.206, NS) or $Dab1^{lacZ/lacZ}$ mice (t-test, p = 0.800, NS). Interestingly, comparison of the total call counts between P7 and P14 time points by genotype reveals a differential rate of age-related decline (Figure A3-1D; two-way ANOVA; age effect p = 0.0001, genotype effect p = 0.006, interaction p = 0.0143). The $Dab1^{+/+}$ mice exhibit

a steep fall-off in calling amount between P7 and P14 (t-test, $p < 0.0001$) in line with the normal developmental decline of isolation calling [245]. Comparatively, in Dab1$^{+/lacZ}$ mice the decline from P7 to P14 was less significant (t-test, $p = 0.001$); and the Dab1$^{lacZ/lacZ}$ mice did not significantly differ in the amount of calling between P7 and P14 (t-test, $p = 0.275$). These findings indicate altered vocal developmental trajectories for the Dab1 reduced (Dab1$^{+/lacZ}$) and null (Dab1$^{lacZ/lacZ}$) mice.

Dab1 genotype affects P7 call repertoires

Next, the types of calls were analyzed to determine any genotype-dependent differences (Figure A3-2A). In addition to an altered amount of calling, described above, the types of calls were also altered in a gene dose-dependent manner (Figures A3-2B, A3-2C). As with call number, the call repertoire exhibited a great deal of variability between pups, even within the same genotype. To enable comparison, data were normalized to create pie charts depicting the combined call repertoire for each genotype (See Methods; Figure A3-B). At P7, *Dab1*$^{+/+}$ mice (N = 13,345 calls from 21 mice) had a relatively diverse repertoire. When comparing calls from *Dab1*$^{+/+}$ and *Dab1*$^{+/lacZ}$ pups (N = 11,075 calls, from 22 mice), *Dab1*$^{+/lacZ}$ had significantly more upward call types. Otherwise, *Dab1*$^{+/lacZ}$ mice had an intermediate phenotype. Trends that placed the *Dab1*$^{+/lacZ}$ between the *Dab1*$^{+/+}$ and *Dab1*$^{lacZ/lacZ}$ mice include an intermediate level of the downward and frequency step call types. The *Dab1*$^{lacZ/lacZ}$ pups (N = 1,317 calls, from 7 mice) exhibited a relatively restricted repertoire comprised of significantly more short and downward calls than found in the other genotypes. The *Dab1*$^{+/+}$ pups made significantly more complex and frequency step calls than did the *Dab1*$^{lacZ/lacZ}$ mice. *Dab1*$^{lacZ/lacZ}$ pups also had significantly more of an

unusual call type, the triple, than the other groups. Chevron, flat, harmonic, two-syllable, double, and miscellaneous call types did not differ significantly based on genotype. A few sex differences in repertoire at P7 were found for *Dab1*$^{lacZ/lacZ}$. However, based on the lack of sex effect on repertoire in *Dab1*$^{+/lacZ}$, call type data were pooled across sexes to provide greater power for further analysis. Notably, repertoire analysis of all wild-type mice across experiments pooled revealed no repertoire differences based on sex.

At P14, Dab1$^{+/+}$ and Dab1$^{+/lacZ}$ repertoires were fairly similar, while those of Dab1$^{lacZ/lacZ}$ mice appeared more restricted than Dab1$^{+/+}$ (Figure A3-3A). There were however, no statistically significant differences in call repertoires (Figure A3-3B; syllables analyzed: Dab1$^{+/+}$ N = 1,141 from 13 mice, Dab1$^{+/lacZ}$ N = 2,518 from 13 mice, Dab1$^{lacZ/lacZ}$ N = 684 from 7 mice). The lack of statistical significance is likely due to the relatively low numbers of calls made at this time point, especially by Dab1$^{lacZ/lacZ}$ mice. In summary, at P7, Dab1$^{+/lacZ}$ and Dab1$^{lacZ/lacZ}$ mice exhibited partially and extremely restricted call repertoires, respectively, relative to the Dab1$^{+/+}$ mice. Dab1$^{lacZ/lacZ}$ repertoires included a decreasing level of some of the more elaborate call types and an increase in some of the simpler ones. Notably, despite gross motor deficits, Dab1$^{lacZ/lacZ}$ pups were able to make a majority of the call types described. Thus, changes in their repertoires may not be fully attributable to global motor deficits.

Effect of Vldlr ablation on calling rates at P7 and P14

Vldlr and Apoer2 are high-affinity Reelin receptors essential for transduction of the signal to Dab1 (Figure A3-4A). To further test that vocal deficits could be related to Vldlr insufficiency, we examined the effect of Vldlr deletion with or without Apoer2. Wild-type ($Vldlr^{+/+}/Apoer2^{+/+}$; N = 12), Vldlr single receptor mutants ($Vldlr^{-/-}/Apoer2^{+/+}$; N = 18), and Vldlr/Apoer2 double receptor mutants ($Vldlr^{-/-}/Apoer2^{-/-}$; N = 4) were recorded at P7 and P14. At P7, there were no significant differences in the number of calls emitted by each group and no effect of sex (Figure A3-4B; two-way ANOVA, genotype effect $p = 0.226$, sex effect $p = 0.447$, interaction $p = 0.700$, NS). This lack of genotype effect on call amount at P7 could be due to the low number of double mutants obtained for recording (N = 4). Upon closer examination, the call counts of $Vldlr^{-/-}/Apoer2^{-/-}$ mice (mean call count = 736.5) were close to statistical significance as being lower than those of the $Vldlr^{-/-}/Apoer2^{+/+}$ mice (mean call count = 1504; t-test $p = 0.057$).

At P14, there were no differences in calling amount based on the Vldlror Apoer2 genotype, but there was a sex difference (two-way ANOVA, genotype effect $p = 0.325$, sex effect $p = 0.010$, interaction $p = 0.224$). The males appear to be more adversely affected and called less than the females. Developmental trajectories of each group were then examined (two-way ANOVA; genotype effect $p = 0.220$, age effect $p = 0.001$, interaction $p = 0.085$). There was a significant decrease in calling from P7 to P14 (Figure A3-4D) by the $Vldlr^{+/+}/Apoer2^{+/+}$ (t-test $p = 0.0001$) and $Vldlr^{-/-}/Apoer2^{+/+}$ mice (t-test $p < 0.0001$). $Vldlr^{-/-}/Apoer2^{-/-}$ mice, much like the $Dab1^{lacZ/lacZ}$ mice, did not have a significant difference in call counts between P7 and P14 (t-test $p = 0.459$,NS). The normal developmental decline in calling rate was observed for both $Vldr^{+/+}/Apoer2^{+/+}$ and $Vldlr^{-/-}/Apoer2^{+/+}$ mice and but not for $Vldlr^{-/-}/Apoer2^{-/-}$ mice.

Vldlr/Apoer2 genotype affects call repertoires at P7 and P14

The call repertoire based on presence of Vldlr was then assessed at P7 and P14 (Figures A3-5, A3-6). Despite the high degree of individual variability, there were significant differences in call usage that paralleled what was observed for Dab1 mice (Figure A3-5B). To improve power, data were pooled across sex, as only minimal sex differences were observed. Overall, the severity of receptor deficiency was inversely related to the call repertoire, with greater deficiencies corresponding to more restricted repertoires (Figure A3-5A). At P7, the $Vldlr^{+/+}/Apoer2^{+/+}$ pups (N = 15046 calls) made more frequency step calls and fewer short calls compared to the other groups. There were significantly more short calls and fewer frequency step calls in both $Vldlr^{-/-}/Apoer2^{+/+}$ (N = 25,522 calls) and $Vldlr^{-/-}/Apoer2^{-/-}$ mice (N = 2,946 calls) compared to $Vldlr^{+/+}/Apoer2^{+/+}$ mice (Figure A3-5B). There was a significant increase in the upward call type in the P7 $Vldlr^{-/-}/Apoer2^{+/+}$ group only, which parallels an increase in upward call type observed in $Dab1^{+/lacZ}$ pups. No significant differences were found for the other call types. Thus, at P7 the $Vldlr^{-/-}/Apoer2^{+/+}$ pups had altered calling behavior reminiscent of that observed in $Dab1^{+/lacZ}$heterozygotes; and extremely restricted repertoires were observed in both $Vldlr^{-/-}/Apoer2^{-/-}$ and $Dab1^{lacZ/lacZ}$ pups.

At P14, repertoire analysis revealed significant differences based on Vldlr genotype (Figure A3-6; $Vldlr^{+/+}/Apoer2^{+/+}$, N = 1,021 calls, $Vldlr^{-/-}/Apoer2^{+/+}$, N = 3,360 calls, $Vldlr^{-/-}/Apoer2^{-/-}$, N = 1,554 calls). $Vldlr^{+/+}/Apoer2^{+/+}$ mice emitted the double call type significantly less often than $Vldlr^{-/-}/Apoer2^{+/+}$ and $Vldlr^{-/-}/Apoer2^{-/-}$ mice, and the short call type less than $Vldlr^{-/-}/Apoer2^{-/-}$ mice. Miscellaneous call types were significantly expanded in

the Vldlr$^{-/-}$/Apoer2$^{-/-}$ compared to the Vldlr$^{-/-}$/Apoer2$^{+/+}$ pups. These findings reflect an extremely restricted repertoire of the Vldlr$^{-/-}$/Apoer2$^{-/-}$ at P14 as was shown in these animals at P7. This was also true of the single receptor mutants, albeit to a lesser degree. Thus, absence of Vldlr, and Vldlr with Apoer2, had a significant effect on calling repertoire at P7 and P14.

Parallel effects of Dab1 and Vldlr genotypes on repertoire correlation and syntax similarity

Repertoire correlation analysis of all pups was performed at P7 (Figure A3-7A,B) to provide a measure of how similar individual repertories are within a given genotype. A high correlation between repertoires (positive correlation values, denoted by red) indicates convergence on similar call type usage, while a low correlation (negative correlation values, denoted in blue) are indicative of very different call usage between individuals (Figure A3-7C,D). Overall, Dab1 mice exhibited a gene-dose dependent effect on repertoire correlation (one-way ANOVA, $p < 0.001$). Pups with the Dab1$^{+/+}$ genotype had the lowest repertoire correlation (average $\rho = 0.22$), followed by Dab1$^{+/lacZ}$ (average $\rho = 0.33$, t-test $p < 0.001$), then Dab1$^{lacZ/lacZ}$ with the highest (average $\rho = 0.50$, t-test, $p = 0.005$). The same was true for pups of the Vldlr/Apoer2genotype (one-way ANOVA, $p = 0.015$) with Vldlr$^{+/+}$/Apoer2$^{+/+}$ having lowest correlation scores (average $\rho = 0.52$), followed by the Vldlr$^{-/-}$/Apoer2$^{+/+}$ (average $\rho = 0.61$, t-test, $p = 0.013$)); and then Vldlr$^{-/-}$/Apoer2$^{-/-}$ (average $\rho = 0.74$, $p = 0.185$, not significant). Repertoire correlations of Vldlr$^{-/-}$/Apoer2$^{-/-}$ pups were significantly lower than those of Vldlr$^{+/+}$/Apoer2$^{+/+}$ pups (t-test $p = 0.038$). These findings indicate an association between highly similar repertoires within the groups of low or no Reelin-signaling pathway components, i.e. Dab1,

Vldlr, and Apoer2. Convergence on similar call types within a genotype would explain increasing repertoire correlation, and was most striking in the Dab1$^{lacZ/lacZ}$ and Vldlr$^{-/-}$/Apoer2$^{-/-}$ pups.

The effect of genotype on call sequence, or syntax, was then examined using syntax similarity analysis of isolation calls for P7 pups (Figure A3-8). This type of analysis shows how alike call transitions are between animals within a given group. High syntax similarity indicates similar types of transitions within a group. There was a significant effect of Dab1genotype on syntax similarity (Figure A3-8A, one-way ANOVA, $p = 0.014$). Syntax similarity was highest for the Dab1$^{+/lacZ}$ mice (Syntax similarity average, $SS = 0.19$) compared to Dab1$^{lacZ/lacZ}$ ($SS = 0.17$) and Dab1$^{+/+}$ pups ($SS = 0.16$). Surprisingly, the Dab1$^{lacZ/lacZ}$ pups had SS scores that were almost identical to that of Dab1$^{+/+}$ pups. There was an effect of Vldlrgenotype on SS as well (Figure A3-8B, one-way ANOVA $p = 0.002$). Vldlr$^{-/-}$/Apoer2$^{+/+}$ pups had higher similarity ($SS = 0.37$) than Vldlr$^{+/+}$/Apoer2$^{+/+}$ pups ($SS = 0.28$) and Vldlr$^{-/-}$/Apoer2$^{-/-}$ pups ($SS = 0.33$). In summary, parallel patterns of syntax similarity were observed across both the Dab1 and Vldlr/Apoer2 mouse lines.

Discussion

Altered isolation vocalizations are a hallmark of ASD-like early phenotype in mice [244]. In order to determine if *Dab1* or *Vldlr* insufficiency impacts patterns of early social communication, we characterized the age related calling patterns in *Dab1* and *Vldlr/Apoer2* deficient mice, generating novel findings. Additionally, we queried whether or not the canonical Reelin-signaling pathway may be responsible for the changes in vocalization seen in $Reln^{+/-}$ and $Reln^{-/-}$ pups [265,270]. Despite extreme inter-individual variation in calling, we found that the *Dab1* genotype profoundly affected the calling rate and repertoire of P7 pups in a gene-dose dependent manner. The effect subsided at P14 in concert with the typical overall decrease in calling amount. Our findings reflect an ASD-like communicative pattern: reduced calling amount, reduced variety in syllable usage, and parallel changes seen in $Reln^{+/-}$ and $Reln^{-/-}$ pups.

We examined the effect of *Vldlr* deficiency on vocal phenotype, based on our previous findings which highlighted *Vldlr* as being vocally regulated in the basal ganglia of adult male zebra finches [51]. Changes in other genes that are critical for birdsong learning, including *Cntnp2* and *FoxP2*, have produced abnormal vocal communication patterns in neonatal mice [69,83,250,251,276,277]. These findings underscore shared mechanisms between vocal learning and non-learning species, and validate a cross-species approach. We found that $Vldlr^{-/-}$ genotype alone did not affect calling rate in mice, but did significantly affect call repertoire at both time points. These changes were observed in both $Vldlr^{-/-}/Apoer2^{+/+}$ and $Vldlr^{-/-}/Apoer2^{-/-}$ groups indicating that loss of *Vldlr* is sufficient to produce these changes to the vocal repertoire. This

limited syllable usage reflects a subtle ASD-like phenotype. The early vocal behavior of *Vldlr* insufficient pups had not been previously characterized.

Both *Dab1* and *Vldlr* gene dose affected the diversity of call repertoires, resulting in simpler call types (no frequency modulation, short duration) with fewer elaborate calls (jump containing, harmonic stacking, long duration). Parallels between the two mouse lines are further underscored by similarities in both the repertoire correlation and the syntax similarity measures. The more repetitive or stereotyped sequencing in both the $Dab1^{+/lacZ}$ and $Vldlr^{-/-}/Apoer2^{+/+}$lines may reflect a subtle vocal phenotype not uncovered by call count and repertoire analyses. The genetic changes in these lines are very different and thus convergence on a high degree of syntax similarity was not predicted. In auditory playback experiments, adult female mice prefer greater call complexity from both adult males and neonates [247,278]. It would therefore be advantageous for pups to emit more elaborate call types in order to be retrieved and thus survive. The restricted repertoire and convergence on simple syllable usage seen here in *Dab1* and *Vldlr* deficient pups would thus be maladaptive, as is the reduction in calling rate as dams prefer to retrieve pups that call more [246].

Sex is another factor contributing to ASD etiology. Because ASD is more prevalent in males, we characterized early vocal phenotypes in each sex and compared them, expecting an exacerbated phenotype in males. To our surprise, when pooling across wild-type controls of both lines, there was no sex difference in calling rate or repertoire. Some minimal sex differences in repertoire were observed in *Dab1* and *Vldlr* deficient pups at P7, but none that suggested that one sex was more adversely affected by the gene loss of *Dab1* or *Vldlr* than the other. Thus, sex does

not appear to interact with *Dab1* or *Vldlr/Apoer2*genotype to produce a more pronounced vocal phenotype. Prior studies provide conflicting reports regarding sex differences in the calling behavior of rodents with some indicating that male neonatal rats and mice call more, or that female mice do, or that there is no difference [279]. These disparate findings indicate that each species and strain should be individually tested rather than generalizing between studies regarding the influence of sex on vocal communication.

Loss of neonatal call type diversity is associated with reduced Reelin signaling as demonstrated here and in prior work. *Reln*$^{+/-}$ *and Reln*$^{-/-}$pups on a similar background as used here, Romano and colleagues [270] observed increased usage of two-syllable call type, and reduced numbers of short and flat call types with increasing *Reelin* insufficiency. In our study, we likewise observe an expansion of some call types and a reduction in others. While the exact call types differed, in both studies, increasingly restricted repertoires emerged in a gene-dose dependent manner. This similar gene-dose restriction across *Reelin, Dab1,* and *Vldlr/Apoer2* lines indicates a newly discovered function of the canonical Reelin-signaling pathway in shaping call-type usage.

ASD is a neurodevelopmental disorder in humans, and diagnosis is based, in part, on altered developmental trajectories and unusual social communication patterns [240]. *Reln*$^{+/-}$ *and Reln*$^{-/-}$ mouse pups exhibit differential vocal developmental trajectories; *Reln*$^{+/-}$ pups have a delayed peak in calling and *Reln*$^{-/-}$ pups lack a peak in calling [270]. We observed similarly altered trajectories for *Dab1*$^{+/lacZ}$*, Dab1*$^{lacZ/lacZ}$ *and Vldlr*$^{-/-}$*Apoer2*$^{-/-}$ mice. These findings also suggest that, like *Reln*$^{+/-}$ mice [255], *Dab1* insufficient mice may serve as a good ASD-risk mouse model.

Future studies could determine whether or not these mice exhibit additional ASD-like behavioral features including repetitive behavior, decreased sociability, or behavioral inflexibility as adults. Once more is understood about the cellular phenotypes underlying Reelin signaling in the basal ganglia, targeted *Dab1* knock-out mice could be used to determine if a vocal phenotype is still present.

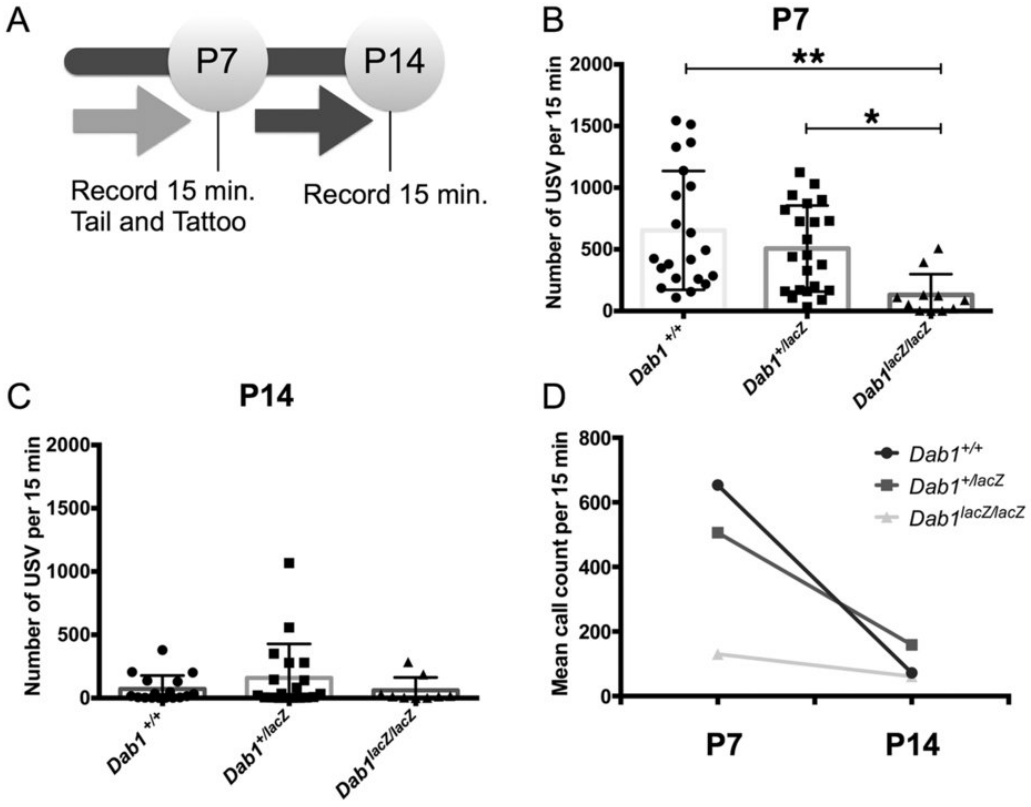
Building on previous work [270], our findings identify a new role for the Reelin-signaling pathway in early vocal phenotypes in mice. It is noteworthy that any differences at all were observed in calling phenotype considering the high degree of inter-individual differences, particularly in call repertoire, that typify these vocalizations. Moreover, mouse pups congenitally engineered to lack a neocortex and hippocampus have indistinguishable calling patterns from wild-type pups [280], emphasizing the significance of the deficits observed here. Since the lack of a cortex does not lead to abnormal calling, the deficits observed here may arise from alterations in subcortical structures. Notably, the basal ganglia has an established role in vocal learning [281], cortico-striatal plasticity is altered in Reelin insufficient mice [258], and abnormal basal ganglia connectivity and excitability are associated with ASD [263,264]. Together these observations provide a relevant yet understudied anatomical locus for future determination of the Reelin-associated neurodevelopmental mechanisms behind early vocal phenotypes.

Acknowledgements

We thank Ava DeLu for help in processing USVs, and Aly Mulji for assistance with mouse breeding and care. This work was supported by the NIH T32 HD07228 (ERF), NIH R01 MH070712 (SAW), NSF IOB-0924143 (PEP).

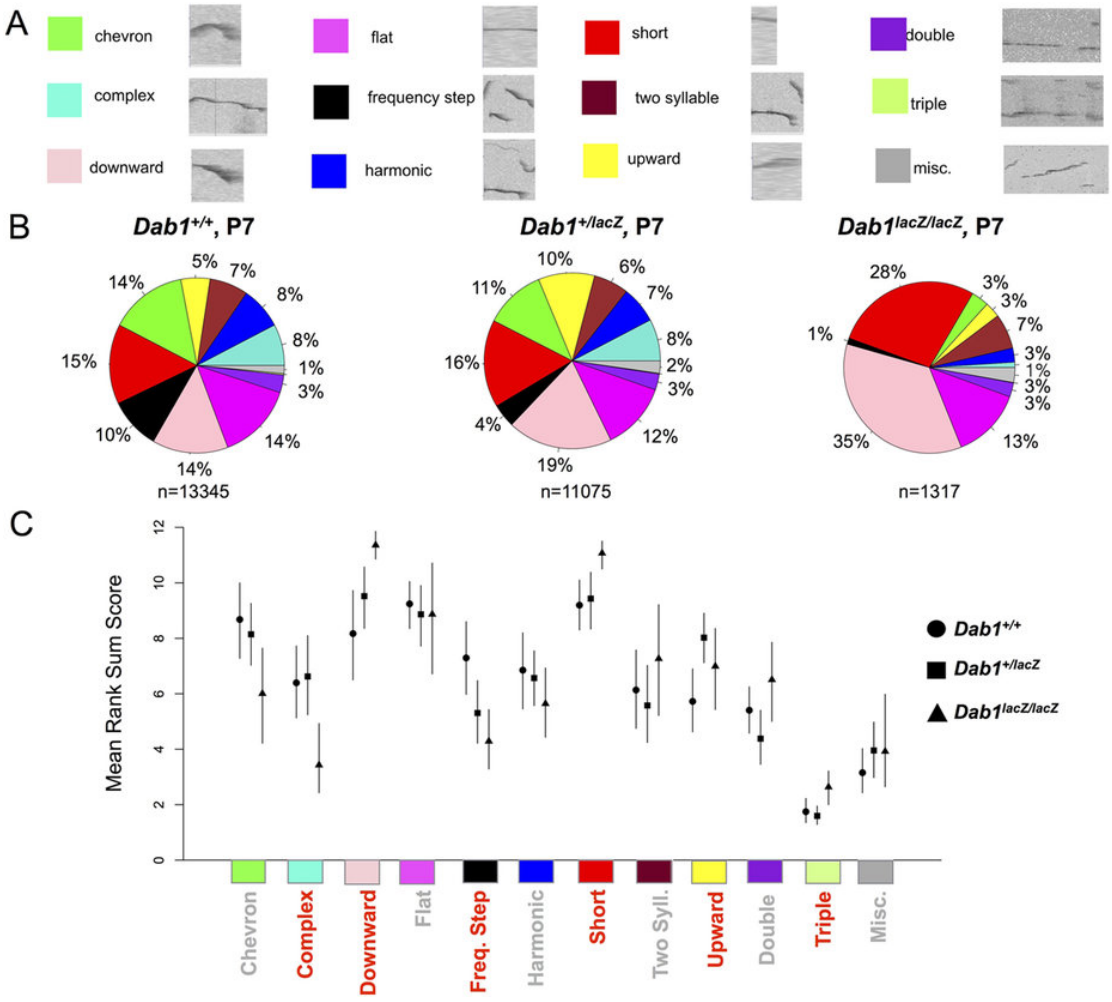
Figures

Figure A3-1: *Dab1* genotype and postnatal age affect pup isolation call amounts.



a) Experimental paradigm. **b)** At P7, *Dab1* mice exhibit gene-dose dependent effects on call number: wild-type pups ($Dab1^{+/+}$) call the most, followed by heterozygote pups ($Dab1^{+/lacZ}$), while homozygous mutant pups ($Dab1^{lacZ/lacZ}$) call the least (**p = 0.0001; *p = 0.0002). **c)** At P14, no differences between call rates are observed. **d)** Developmental trajectories in calling amount vary by genotype. Between P7 and P14, $Dab1^{+/+}$ pups exhibit the steepest decline (**p < 0.0001) followed by $Dab1^{+/lacZ}$ pups (*p = 0.0007). Call rate did not decline in $Dab1^{lacZ/lacZ}$ mice (p = 0.297, NS).

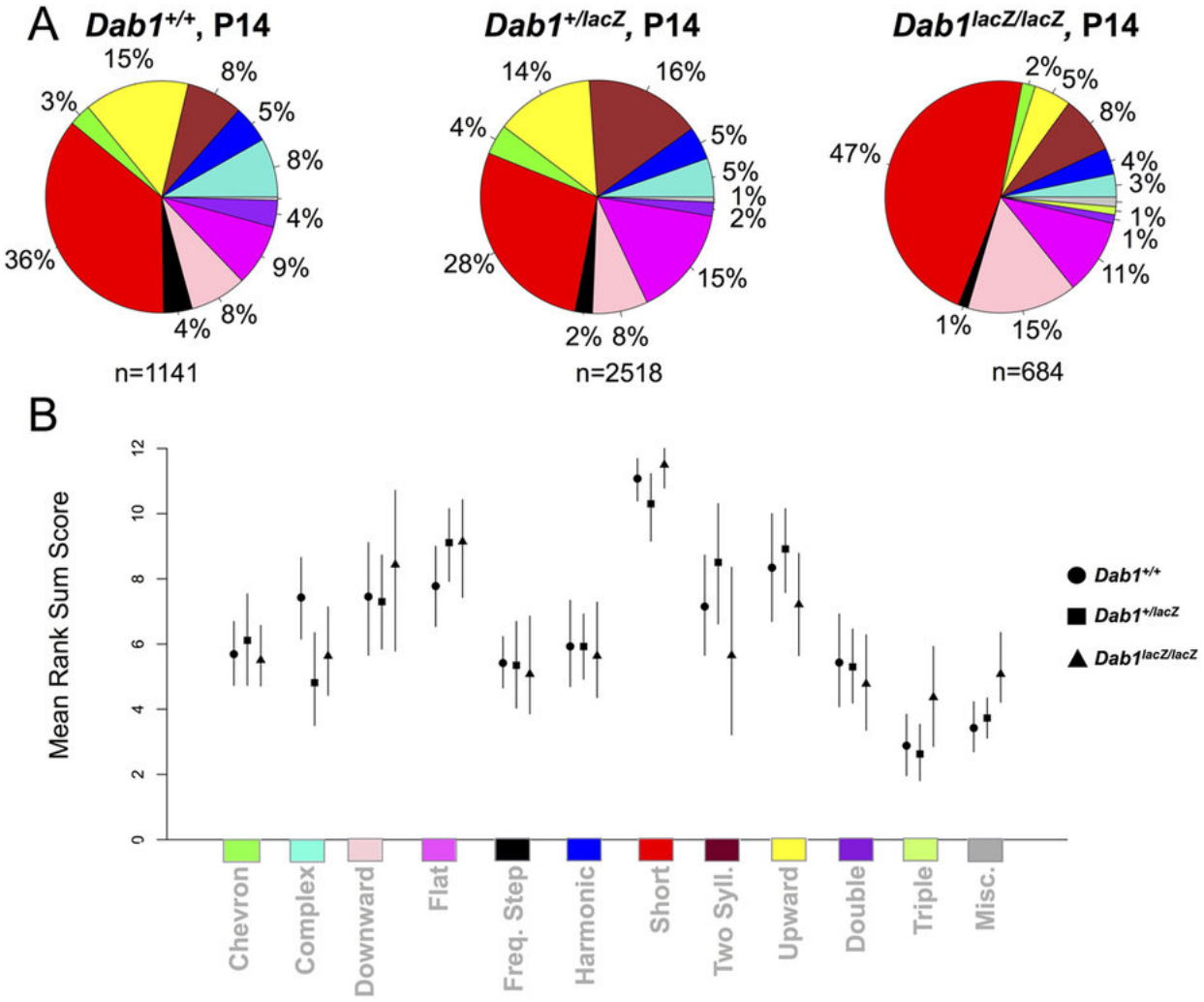
Figure A3-2: *Dab1* genotype affects P7 call repertoire.



a) Representative syllables for each call cluster, known as eigen syllables, are shown with their classifications and representative colors. (The same colors are used in Figures A4-3, A4-5 and A4-6) **b)** Pie charts depict P7 calling repertoires of *Dab1*$^{+/+}$ (N = 13,345 calls from 21 pups), *Dab1*$^{+/lacZ}$ (N = 11,075 calls from 22 pups), and *Dab1*$^{lacZ/lacZ}$ mice (N = 1,317 calls from 7 pups). **c)** Quantitative repertoire analysis. Data are rank sum transformed such that 12 on the y axis denotes high call use probability and 1, low call use probability. Lines indicate 95% confidence intervals, shapes correspond to genotypes: Circle (*Dab1*$^{+/+}$), square (*Dab1*$^{+/lacZ}$), and triangle (*Dab1*$^{lacZ/lacZ}$).

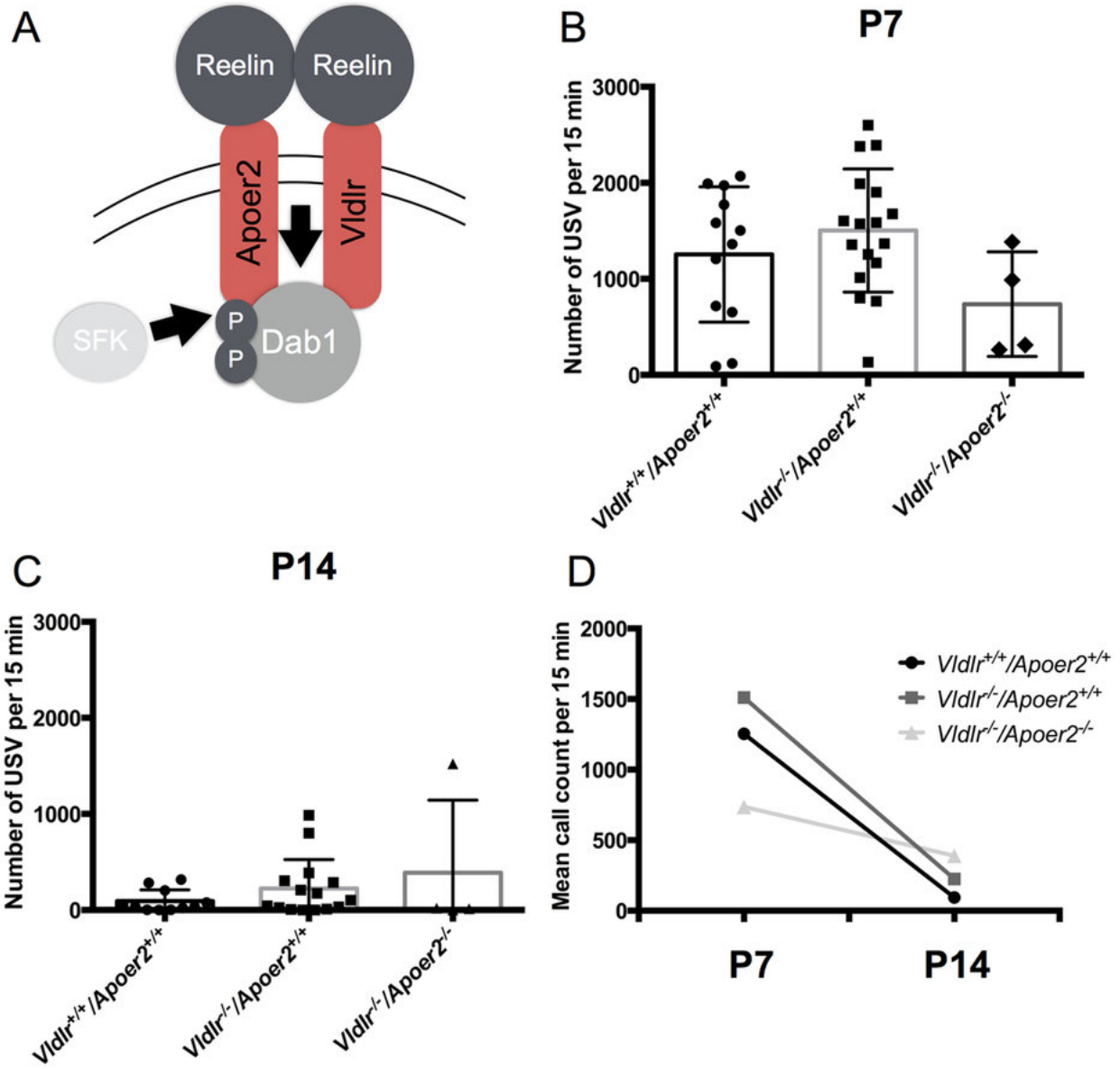
Significant differences are indicated when call categories are highlighted in red on the x axis, and the 95% confidence intervals do not overlap between one or more genotypes. Differences are found for the following categories: complex, downward, frequency step, short, upward, and triple.

Figure A3-3: *Dab1* genotype does not affect P14 call repertoire



a) Call repertoires for wild-type (N = 1,141 calls from 11 pups), heterozygous (N = 2,518 calls from 19 pups), and homozygous mice (N = 684 calls from 3 pups). **b)** For each genotype, quantification shows 95% confidence intervals resampled about the median call usage. Shapes correspond to genotypes: circle (*Dab1*$^{+/+}$), square (*Dab1*$^{+/lacZ}$), and triangle (*Dab1*$^{lacZ/lacZ}$). There are no significant differences.

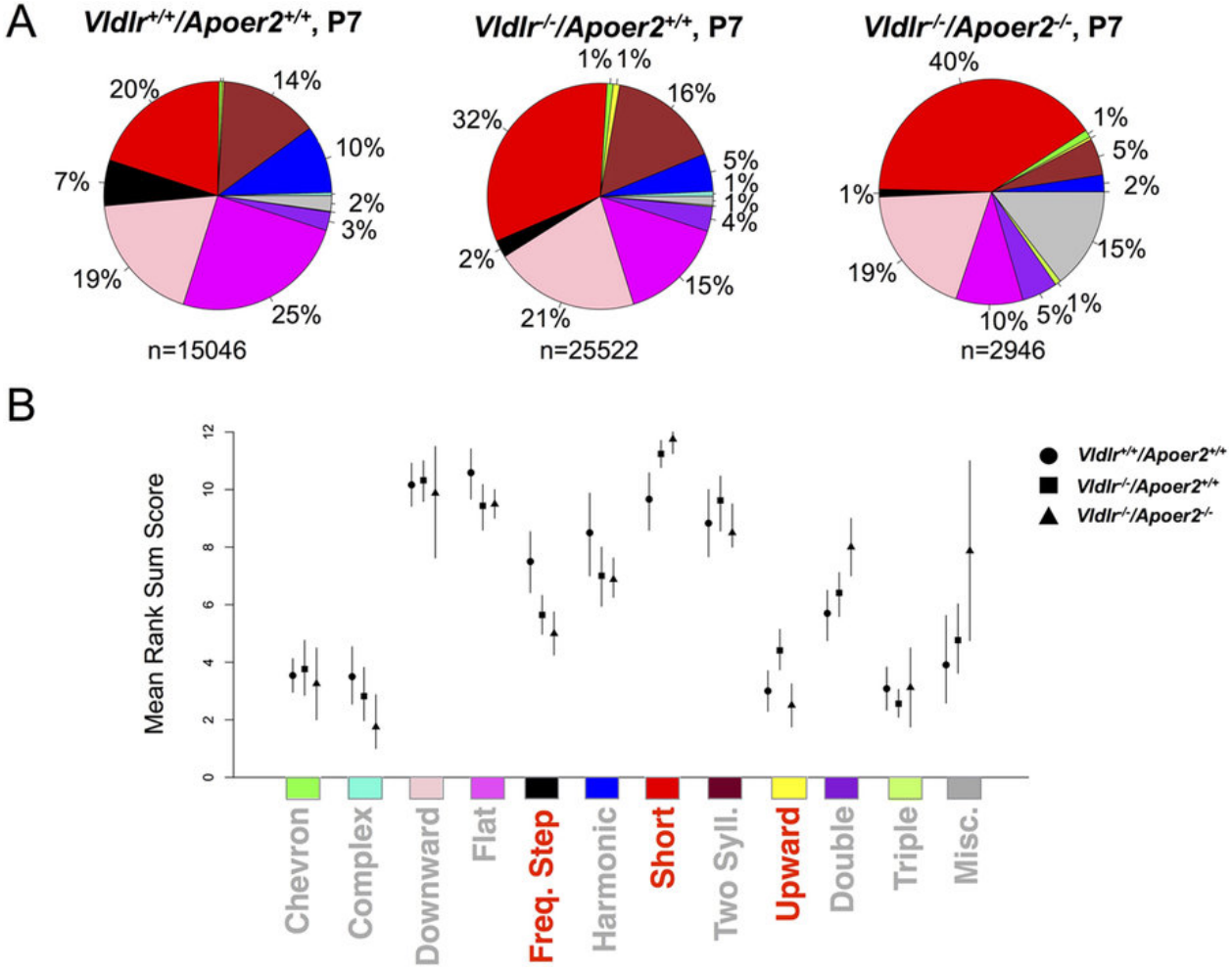
Figure A3-4: *Vldlr* and *Vldlr/Apoer2* insufficient pups have altered developmental trajectories in calling amount.



a) Schematic depicts the canonical Reelin-signaling pathway. Signal is transduced via Reelin binding to receptors *Vldlr* and *Apoer2* to initiate phosphorylation of *Dab1* via Src-family kinases. **b)** Quantification of P7 call counts from wild-type pups ($Vldlr^{+/+}/Apoer2^{+/+}$, N = 12), *Vldlr* single receptor mutants ($Vldlr^{-/-}/Apoer2^{+/+}$; N = 18) and double receptor mutants ($Vldlr^{-/-}/Apoer2^{-/-}$;

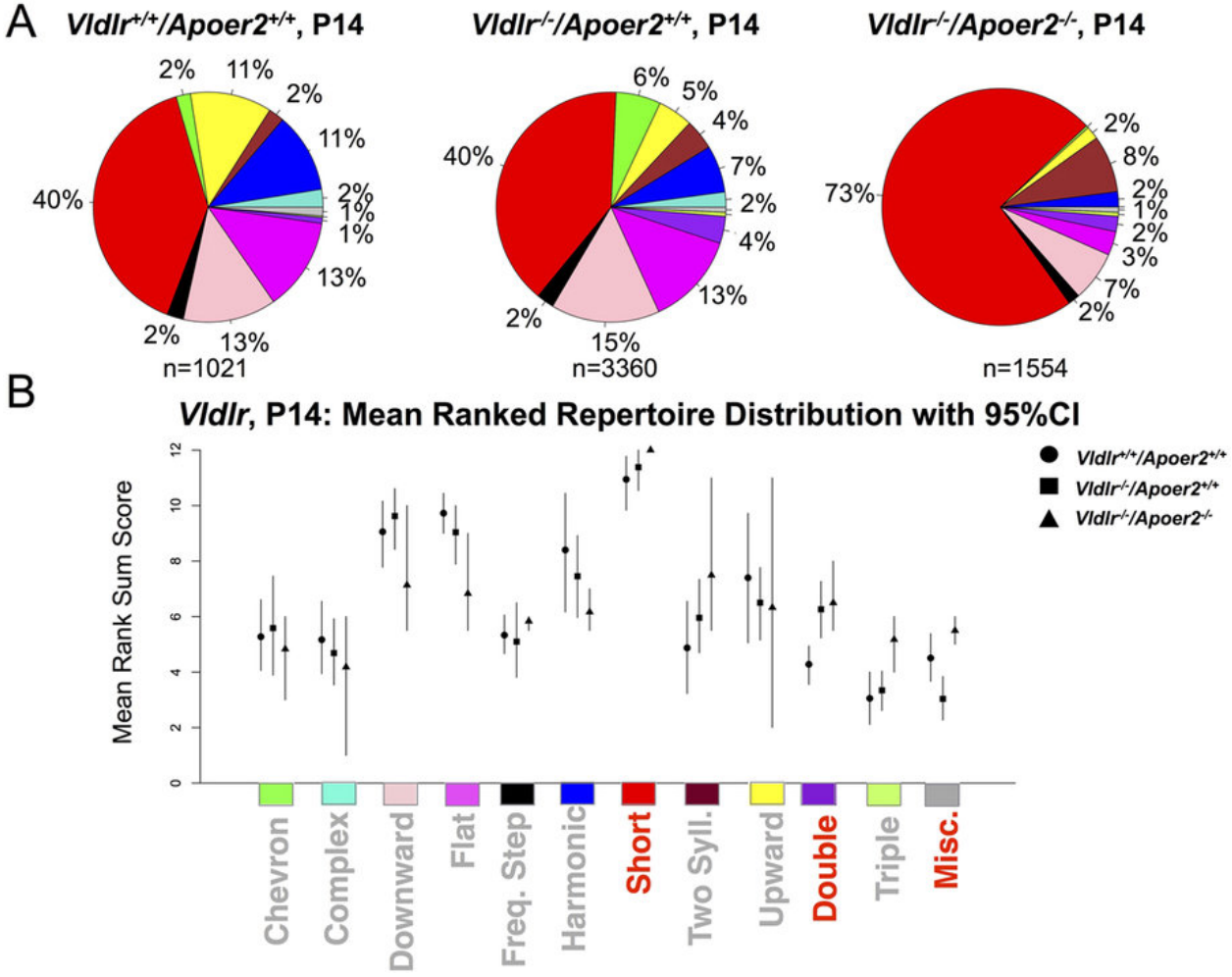
N = 4). Trends suggest that the double receptor mutants call less than the other genotypes. **c)** Quantification of P14 call counts for pups of all three genotypes reveal no significant differences. **d)** Developmental trajectories between P7 and P14 differ by genotype. Call amounts of $Vldlr^{+/+}/Apoer2^{+/+}$ pups ($p = 0.0001$) and $Vldlr^{-/-}/Apoer2^{+/+}$ pups decline (**$p < 0.0001$) but those of $Vldlr^{-/-}/Apoer2^{-/-}$ pups do not ($p = 0.486$, NS).

Figure A3-5: P7 call repertoire is influenced by *Vldlr*/*Apoer2* genotype.



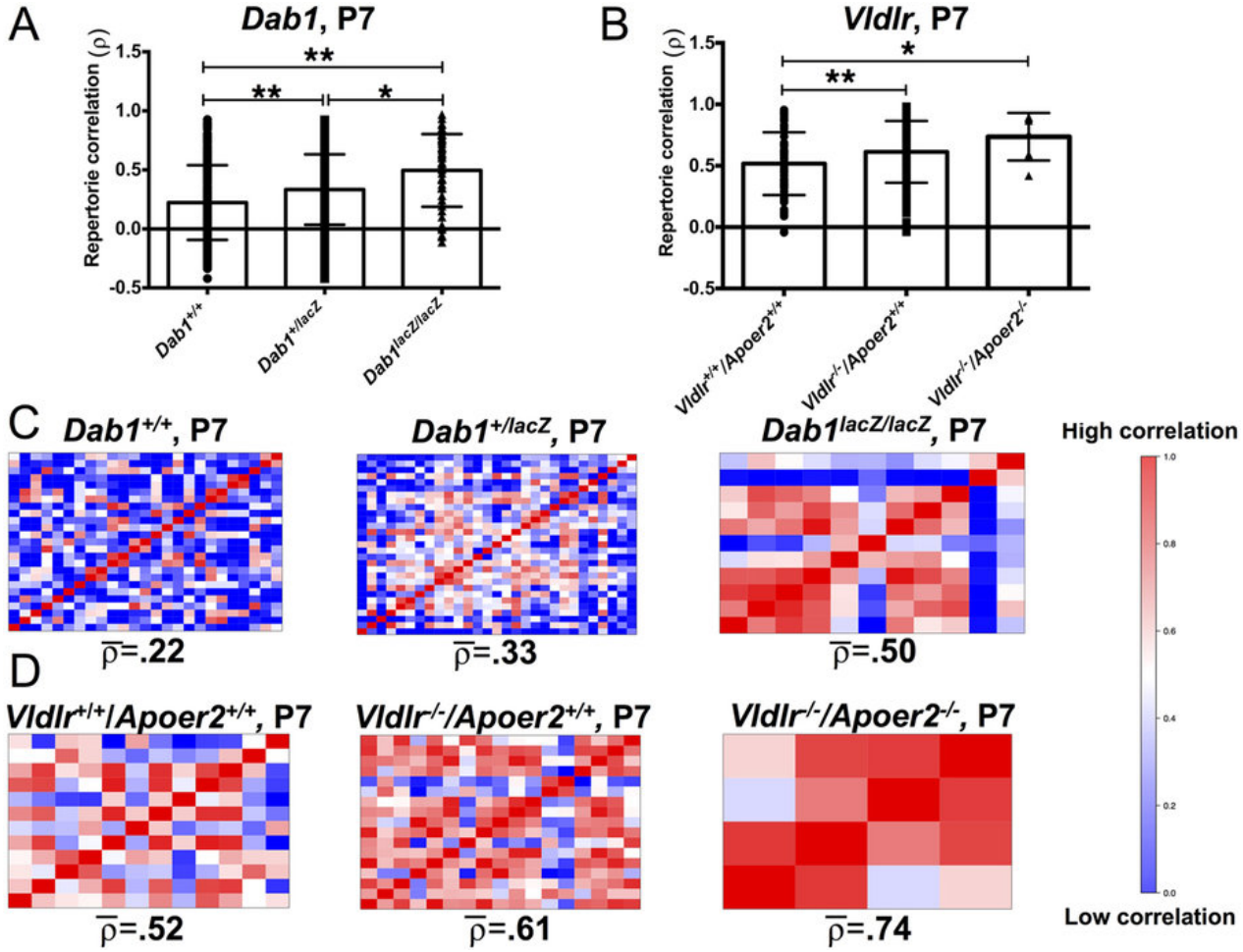
a) Repertoires were determined and are depicted as in Figure A3-2. Mice of *Vldlr*$^{+/+}$/*Apoer2*$^{+/+}$ (N = 15,046 calls from 12 pups), *Vldlr*$^{-/-}$/*Apoer2*$^{+/+}$ (N = 25,522 calls from 18 pups), and *Vldlr*$^{-/-}$/*Apoer2*$^{-/-}$ genotypes (N = 2,946 from 4 pups) exhibit increasingly restricted repertoires, respectively. **b)** Quantification of calling repertoire differences between genotypes. Each shape corresponds to a genotype: Circle signifies *Vldlr*$^{+/+}$/*Apoer2*$^{+/+}$; square *Vldlr*$^{-/-}$/*Apoer2*$^{+/+}$, triangle *Vldlr*$^{-/-}$/*Apoer2*$^{-/-}$. Significant differences (highlighted in red) were found for the following call types: frequency steps, short and upward.

Figure A3-6: P14 call repertoire is influenced by *Vldlr*/*Apoer2* genotype.



(**A**) Pie charts depict call repertoires of *Vldlr*$^{+/+}$/*Apoer2*$^{+/+}$ (N = 1,021 calls from 9 pups), *Vldlr*$^{-/-}$/*Apoer2*$^{+/+}$ (N = 3,360 from 13 pups) and *Vldlr*$^{-/-}$/*Apoer2*$^{-/-}$ mice (N = 1,554 calls from 3 pups). (**B**) Repertoire analysis shows significant differences in the short, double and miscellaneous categories as revealed by non-overlapping confidence intervals.

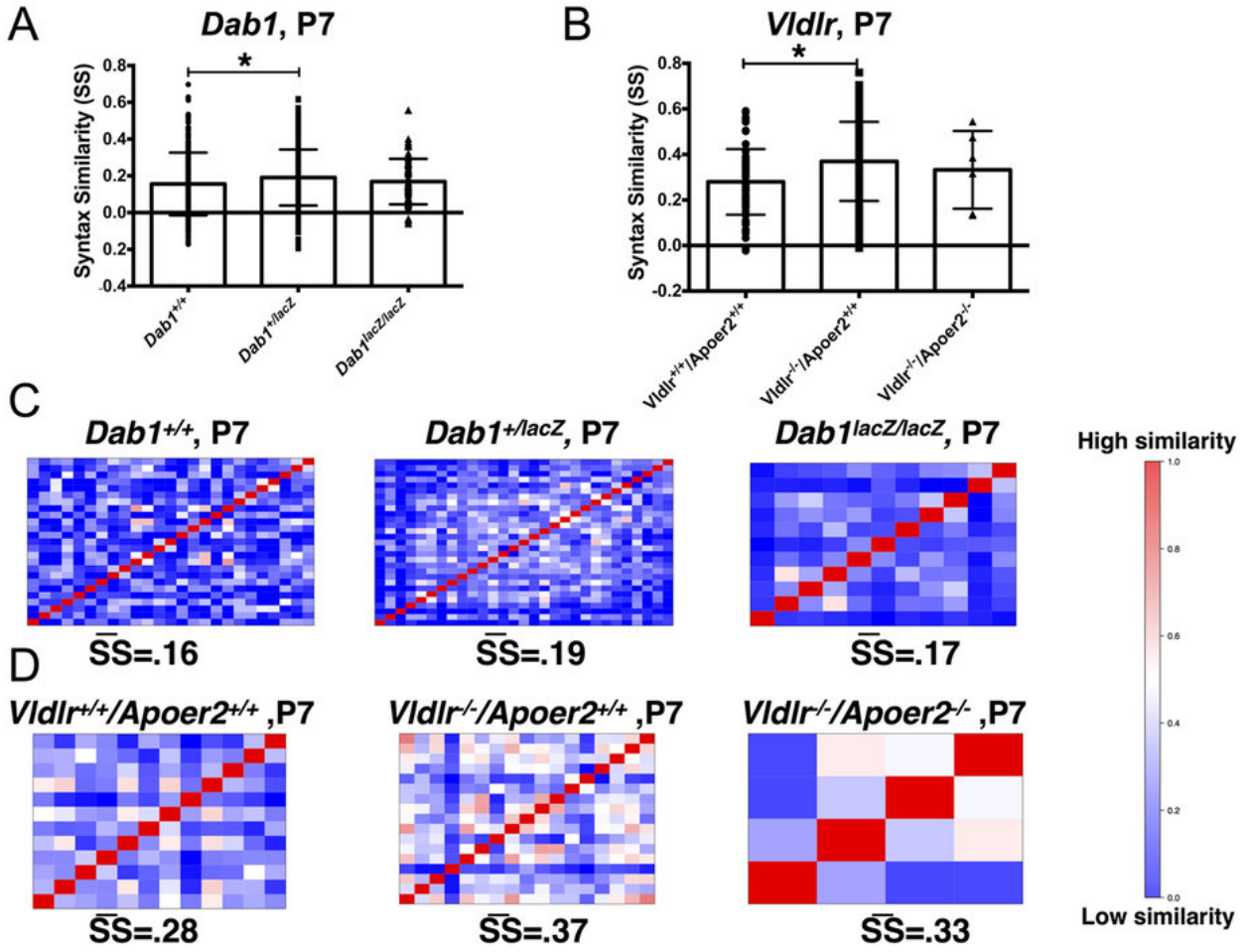
Figure A3-7: *Dab1* and *Vldlr*/*Apoer2* pups exhibit a gene-dose dependent increase in repertoire correlation at P7.



(**a,b**) Repertoire correlation scores across all animals of each genotype. $Dab1^{+/lacZ}$ pups have higher scores, reflecting a more restricted repertoire than $Dab1^{+/+}$ pups (**$p < 0.0001$). $Dab1^{lacZ/lacZ}$ have higher scores than either $Dab1^{+/lacZ}$ (*$p = 0.0005$) or $Dab1^{+/+}$ pups (**$p < 0.0001$). $Vldlr^{-/-}/Apoer2^{+/+}$ have higher scores than $Vldlr^{+/+}/Apoer2^{+/+}$ (**$p = 0.0131$), and $Vldlr^{-/-}/Apoer2^{-/-}$ pups exhibit a higher repertoire correlation than $Vldlr^{+/+}/Apoer2^{+/+}$ pups (*$p = 0.038$). This pattern of increasing repertoire correlation in gene reduced or deficient pups is parallel across both lines (*Dab1, Vldlr/Apoer2*). (**c,d**) Repertoire correlation matrices

for *Dab1* and *Vldlr/Apoer2* mice. Average repertoire correlation score is shown below each matrix (rho). Red indicates high correlation, and blue indicates low correlation on a scale of 0–1.

Figure A3-8: *Dab1*$^{+/lacZ}$ and *Vldlr*$^{-/-}$/*Apoer2*$^{+/+}$ pups have high syntax similarity scores.



a,b) Syntax similarity scores across all animals of each genotype. *Dab1*$^{+/lacZ}$ pups have higher scores than *Dab1*$^{+/+}$ pups (*p = 0.005). There is no detectable difference between *Dab1*$^{lacZ/lacZ}$ and *Dab1*$^{+/+}$ pups. *Vldlr*$^{-/-}$/*Apoer2*$^{+/+}$ pups have higher scores than *Vldlr*$^{+/+}$/*Apoer2*$^{+/+}$ (**p = 0.0002). *Vldlr*$^{-/-}$/*Apoer2*$^{-/-}$ mice do not differ from *Vldlr*$^{+/+}$/*Apoer2*$^{+/+}$. **c,d)** Syntax similarity matrices for *Dab1* and *Vldlr/Apoer2* mice. Average repertoire correlation score is shown below each matrix (SS). Red indicates high correlation, and blue indicates low correlation on a scale of 0–1.

References

[1] Clayton DF. The genomics of memory and learning in songbirds. Annu Rev Genomics Hum Genet 2013; 14:45–65. doi:10.1146/annurev-genom-090711-163809.

[2] Simpson HB, Vicario DS. Brain pathways for learned and unlearned vocalizations differ in zebra finches. J Neurosci 1990; 10:1541–56.

[3] Vates GE, Nottebohm F. Feedback circuitry within a song-learning pathway. Proc Natl Acad Sci USA 1995; 92:5139–43.

[4] Goldberg JH, Fee MS. Vocal babbling in songbirds requires the basal ganglia-recipient motor thalamus but not the basal ganglia. J Neurophysiol 2011; 105:2729–39. doi:10.1152/jn.00823.2010.

[5] Kao MH, Brainard MS. Lesions of an avian basal ganglia circuit prevent context-dependent changes to song variability. J Neurophysiol 2006; 96:1441–55. doi:10.1152/jn.01138.2005.

[6] Kao MH, Doupe AJ, Brainard MS. Contributions of an avian basal ganglia-forebrain circuit to real-time modulation of song. Nature 2005; 433:638–43. doi:10.1038/nature03127.

[7] Boettiger CA, Doupe AJ. Intrinsic and thalamic excitatory inputs onto songbird LMAN neurons differ in their pharmacological and temporal properties. J Neurophysiol 1998; 79:2615–28.

[8] Farries MA, Perkel DJ. A telencephalic nucleus essential for song learning contains neurons with physiological characteristics of both striatum and globus pallidus. J Neurosci 2002; 22:3776–87. doi:10.1016/0165-0270(89)90131-3.

[9] Luo M, Perkel DJ. A GABAergic, strongly inhibitory projection to a thalamic nucleus in the zebra finch song system. J Neurosci 1999; 19:6700–11.

[10] Pfenning AR, Hara E, Whitney O, Rivas MV, Wang R, Roulhac PL, Howard JT, Wirthlin M, Lovell PV, Ganapathy G, Mouncastle J, Moseley MA, Thompson JW, Soderblom EJ, Iriki A, Kato M, Gilbert MTP, Zhang G, Bakken T, Bongaarts A, Bernard A, Lein E, Mello CV, Hartemink AJ, Jarvis ED. Convergent transcriptional specializations in the brains of humans and song-learning birds. Science 2014; 346:1256846–6. doi:10.1126/science.1256846.

[11] Lai CS, Fisher SE, Hurst JA, Vargha-Khadem F, Monaco AP. A forkhead-domain gene is mutated in a severe speech and language disorder. Nature 2001; 413:519–23. doi:10.1038/35097076.

[12] Vargha-Khadem F, Watkins K, Alcock K, Fletcher P, Passingham R. Praxic and nonverbal cognitive deficits in a large family with a genetically transmitted speech and language disorder. Proc Natl Acad Sci USA 1995; 92:930–3.

[13] Spiteri E, Konopka G, Coppola G, Bomar J, Oldham M, Ou J, Vernes SC, Fisher SE, Ren B, Geschwind DH. Identification of the transcriptional targets of FOXP2, a gene linked to speech and language, in developing human brain. Am J Hum Genet 2007; 81:1144–57. doi:10.1086/522237.

[14] Vernes SC, Oliver PL, Spiteri E, Lockstone HE, Puliyadi R, Taylor JM, Ho J, Mombereau C, Brewer A, Lowy E, Nicod J, Groszer M, Baban D, Sahgal N, Cazier J-B, Ragoussis J, Davies KE, Geschwind DH, Fisher SE. Foxp2 regulates gene networks implicated in neurite outgrowth in the developing brain. PLoS Genet 2011; 7:e1002145. doi:10.1371/journal.pgen.1002145.

[15] Mello CV, Vicario DS, Clayton DF. Song presentation induces gene expression in the songbird forebrain. Proc Natl Acad Sci USA 1992; 89:6818–22.

[16] Mello C, Nottebohm F, Clayton D. Repeated exposure to one song leads to a rapid and persistent decline in an immediate early gene's response to that song in zebra finch telencephalon. J Neurosci 1995; 15:6919–25.

[17] Margoliash D. Acoustic parameters underlying the responses of song-specific neurons in the white-crowned sparrow. J Neurosci 1983; 3:1039–57.

[18] Doupe AJ, Konishi M. Song-selective auditory circuits in the vocal control system of the zebra finch. Proc Natl Acad Sci USA 1991; 88:11339–43.

[19] Margoliash D, Fortune ES. Temporal and harmonic combination-sensitive neurons in the zebra finch's HVc. J Neurosci 1992; 12:4309–26.

[20] Vicario DS, Yohay KH. Song-selective auditory input to a forebrain vocal control nucleus in the zebra finch. J Neurobiol 1993; 24:488–505. doi:10.1002/neu.480240407.

[21] Volman SF. Development of neural selectivity for birdsong during vocal learning. J Neurosci 1993; 13:4737–47.

[22] Volman SF. Quantitative assessment of song-selectivity in the zebra finch "high vocal center". J Comp Physiol A 1996; 178:849–62.

[23] Doupe AJ. Song- and order-selective neurons in the songbird anterior forebrain and their emergence during vocal development. J Neurosci 1997; 17:1147–67.

[24] Prather JF, Peters S, Nowicki S, Mooney R. Precise auditory-vocal mirroring in neurons for learned vocal communication. Nature 2008; 451:305–10. doi:10.1038/nature06492.

[25] Jarvis ED, Nottebohm F. Motor-driven gene expression. Proc Natl Acad Sci USA 1997; 94:4097–102.

[26] Jin H, Clayton DF. Localized changes in immediate-early gene regulation during sensory and motor learning in zebra finches. Neuron 1997; 19:1049–59.

[27] Kimpo RR, Doupe AJ. FOS is induced by singing in distinct neuronal populations in a motor network. Neuron 1997; 18:315–25.

[28] Velho TAF, Pinaud R, Rodrigues PV, Mello CV. Co-induction of activity-dependent genes in songbirds. Eur J Neurosci 2005; 22:1667–78. doi:10.1111/j.1460-9568.2005.04369.x.

[29] Li XC, Jarvis ED, Alvarez-Borda B, Lim DA, Nottebohm F. A relationship between behavior, neurotrophin expression, and new neuron survival. Proc Natl Acad Sci USA 2000; 97:8584–9. doi:10.1073/pnas.140222497.

[30] Lombardino AJ, Li X-C, Hertel M, Nottebohm F. Replaceable neurons and neurodegenerative disease share depressed UCHL1 levels. Proc Natl Acad Sci USA 2005; 102:8036–41. doi:10.1073/pnas.0503239102.

[31] Teramitsu I, Kudo LC, London SE, Geschwind DH, White SA. Parallel FoxP1 and FoxP2 expression in songbird and human brain predicts functional interaction. J Neurosci 2004; 24:3152–63. doi:10.1523/jneurosci.5589-03.2004.

[32] Teramitsu I, White SA. FoxP2 regulation during undirected singing in adult songbirds. J Neurosci 2006; 26:7390–4. doi:10.1523/jneurosci.1662-06.2006.

[33] Miller JE, Spiteri E, Condro MC, Dosumu-Johnson RT, Geschwind DH, White SA. Birdsong decreases protein levels of FoxP2, a molecule required for human speech. J Neurophysiol 2008; 100:2015–25. doi:10.1152/jn.90415.2008.

[34] Haesler S, Rochefort C, Georgi B, Licznerski P, Osten P, Scharff C. Incomplete and inaccurate vocal imitation after knockdown of FoxP2 in songbird basal ganglia nucleus Area X. PLoS Biol 2007; 5:e321. doi:10.1371/journal.pbio.0050321.

[35] Heston JB, White SA. Behavior-linked FoxP2 regulation enables zebra finch vocal learning. J Neurosci 2015; 35:2885–94. doi:10.1523/jneurosci.3715-14.2015.

[36] Jarvis ED, Smith VA, Wada K, Rivas MV, McElroy M, Smulders TV, Carninci P, Hayashizaki Y, Dietrich F, Wu X, McConnell P, Yu J, Wang PP, Hartemink AJ, Lin S. A framework for integrating the songbird brain. J Comp Physiol a Neuroethol Sens Neural Behav Physiol 2002; 188:961–80. doi:10.1007/s00359-002-0358-y.

[37] Wada K, Howard JT, McConnell P, Whitney O, Lints T, Rivas MV, Horita H, Patterson MA, White SA, Scharff C, Haesler S, Zhao S, Sakaguchi H, Hagiwara M, Shiraki T, Hirozane-Kishikawa T, Skene P, Hayashizaki Y, Carninci P, Jarvis ED. A molecular neuroethological approach for identifying and characterizing a cascade of behaviorally regulated genes. Proc Natl Acad Sci USA 2006; 103:15212–7. doi:10.1073/pnas.0607098103.

[38] Li X, Wang X-J, Tannenhauser J, Podell S, Mukherjee P, Hertel M, Biane J, Masuda S, Nottebohm F, Gaasterland T. Genomic resources for songbird research and their use in characterizing gene expression during brain development. Proc Natl Acad Sci USA 2007; 104:6834–9. doi:10.1073/pnas.0701619104.

[39] Replogle K, Arnold AP, Ball GF, Band M, Bensch S, Brenowitz EA, Dong S, Drnevich J, Ferris M, George JM, Gong G, Hasselquist D, Hernandez AG, Kim R, Lewin HA, Liu L, Lovell PV, Mello CV, Naurin S, Rodriguez-Zas S, Thimmapuram J, Wade J, Clayton DF. The Songbird Neurogenomics (SoNG) Initiative: community-based tools and strategies for study of brain gene function and evolution. BMC Genomics 2008; 9:131. doi:10.1186/1471-2164-9-131.

[40] Whitney O, Pfenning AR, Howard JT, Blatti CA, Liu F, Ward JM, Wang R, Audet J-N, Kellis M, Mukherjee S, Sinha S, Hartemink AJ, West AE, Jarvis ED. Core and region-enriched networks of behaviorally regulated genes and the singing genome. Science 2014; 346:1256780. doi:10.1126/science.1256780.

[41] Warren WC, Clayton DF, Ellegren H, Arnold AP, Hillier LW, Künstner A, Searle S, White S, Vilella AJ, Fairley S, Heger A, Kong L, Ponting CP, Jarvis ED, Mello CV, Minx P, Lovell P, Velho TAF, Ferris M, Balakrishnan CN, Sinha S, Blatti C, London SE, Li Y, Lin Y-C, George J, Sweedler J, Southey B, Gunaratne P, Watson M, Nam K, Backström N, Smeds L, Nabholz B, Itoh Y, Whitney O, Pfenning AR, Howard J, Völker M, Skinner BM, Griffin DK, Ye L, Mclaren WM, Flicek P, Quesada V, Velasco G,

Lopez-Otin C, Puente XS, Olender T, Lancet D, Smit AFA, Hubley R, Konkel MK, Walker JA, Batzer MA, Gu W, Pollock DD, Chen L, Cheng Z, Eichler EE, Stapley J, Slate J, Ekblom R, Birkhead T, Burke T, Burt D, Scharff C, Adam I, Richard H, Sultan M, Soldatov A, Lehrach H, Edwards SV, Yang S-P, Li X, Graves T, Fulton L, Nelson J, Chinwalla A, Hou S, Mardis ER, Wilson RK. The genome of a songbird. Nature 2010; 464:757–62. doi:10.1038/nature08819.

[42] Mattick JS. RNA regulation: a new genetics? Nat Rev Genet 2004; 5:316–23. doi:10.1038/nrg1321.

[43] Shi Z, Luo G, Fu L, Fang Z, Wang X, Li X. miR-9 and miR-140-5p target FoxP2 and are regulated as a function of the social context of singing behavior in zebra finches. J Neurosci 2013; 33:16510–21. doi:10.1523/jneurosci.0838-13.2013.

[44] Zhang B, Horvath S. A general framework for weighted gene co-expression network analysis. Stat Appl Genet Mol Biol 2005; 4:Article17. doi:10.2202/1544-6115.1128.

[45] Langfelder P, Zhang B, Horvath S. Defining clusters from a hierarchical cluster tree: the Dynamic Tree Cut package for R. Bioinformatics 2008; 24:719–20. doi:10.1093/bioinformatics/btm563.

[46] Lee J-H, Gao C, Peng G, Greer C, Ren S, Wang Y, Xiao X. Analysis of transcriptome complexity through RNA sequencing in normal and failing murine hearts. Circ Res 2011; 109:1332–41. doi:10.1161/circresaha.111.249433.

[47] Berg JM, Geschwind DH. Autism genetics: searching for specificity and convergence. Genome Biol 2012; 13:247. doi:10.1186/gb-2012-13-7-247.

[48] Parikshak NN, Luo R, Zhang A, Won H, Lowe JK, Chandran V, Horvath S, Geschwind DH. Integrative functional genomic analyses implicate specific molecular pathways and circuits in autism. Cell 2013; 155:1008–21. doi:10.1016/j.cell.2013.10.031.

[49] Konopka G, Bomar JM, Winden K, Coppola G, Jonsson ZO, Gao F, Peng S, Preuss TM, Wohlschlegel JA, Geschwind DH. Human-specific transcriptional regulation of CNS development genes by FOXP2. Nature 2009; 462:213–7. doi:10.1038/nature08549.

[50] Yang Y, Han L, Yuan Y, Li J, Hei N, Liang H. Gene co-expression network analysis reveals common system-level properties of prognostic genes across cancer types. Nat Commun 2014; 5:3231. doi:10.1038/ncomms4231.

[51] Hilliard AT, Miller JE, Fraley ER, Horvath S, White SA. Molecular microcircuitry

underlies functional specification in a basal ganglia circuit dedicated to vocal learning. Neuron 2012; 73:537–52. doi:10.1016/j.neuron.2012.01.005.

[52] Drnevich J, Replogle KL, Lovell P, Hahn TP, Johnson F, Mast TG, Nordeen E, Nordeen K, Strand C, London SE, Mukai M, Wingfield JC, Arnold AP, Ball GF, Brenowitz EA, Wade J, Mello CV, Clayton DF. Impact of experience-dependent and -independent factors on gene expression in songbird brain. Proc Natl Acad Sci USA 2012; 109:17245–52. doi:10.1073/pnas.1200655109.

[53] Zinzow-Kramer WM, Horton BM, McKee CD, Michaud JM, Tharp GK, Thomas JW, Tuttle EM, Yi S, Maney DL. Genes located in a chromosomal inversion are correlated with territorial song in white-throated sparrows. Genes Brain Behav 2015; 14:641–54. doi:10.1111/gbb.12252.

[54] Mori C, Wada K. Audition-independent vocal crystallization associated with intrinsic developmental gene expression dynamics. J Neurosci 2015; 35:878–89. doi:10.1523/jneurosci.1804-14.2015.

[55] Doupe AJ, Kuhl PK. Birdsong and human speech: common themes and mechanisms. Annu Rev Neurosci 1999; 22:567–631. doi:10.1146/annurev.neuro.22.1.567.

[56] Lemon RN. Descending pathways in motor control. Annu Rev Neurosci 2008; 31:195–218. doi:10.1146/annurev.neuro.31.060407.125547.

[57] Jürgens U. Neural pathways underlying vocal control. Neurosci Biobehav Rev 2002; 26:235–58. doi:10.1016/S0149-7634(01)00068-9.

[58] Arriaga G, Zhou EP, Jarvis ED. Of mice, birds, and men: the mouse ultrasonic song system has some features similar to humans and song-learning birds. PLoS ONE 2012; 7:e46610. doi:10.1371/journal.pone.0046610.

[59] Clark KL, Halay ED, Lai E, Burley SK. Co-crystal structure of the HNF-3/fork head DNA-recognition motif resembles histone H5. Nature 1993; 364:412–20. doi:10.1038/364412a0.

[60] Vargha-Khadem F, Watkins KE, Price CJ, Ashburner J, Alcock KJ, Connelly A, Frackowiak RS, Friston KJ, Pembrey ME, Mishkin M, Gadian DG, Passingham RE. Neural basis of an inherited speech and language disorder. Proc Natl Acad Sci USA 1998; 95:12695–700. doi:10.1073/pnas.95.21.12695.

[61] Vernes SC, Nicod J, Elahi FM, Coventry JA, Kenny N, Coupe A-M, Bird LE, Davies

KE, Fisher SE. Functional genetic analysis of mutations implicated in a human speech and language disorder. Hum Mol Genet 2006; 15:3154–67. doi:10.1093/hmg/ddl392.

[62] Vernes SC, Newbury DF, Abrahams BS, Winchester L, Nicod J, Groszer M, Alarcón M, Oliver PL, Davies KE, Geschwind DH, Monaco AP, Fisher SE. A functional genetic link between distinct developmental language disorders. N Engl J Med 2008; 359:2337–45. doi:10.1056/nejmoa0802828.

[63] Hall MF. Evolutionary aspects of estrildid song. Symp Zool Soc Lond 1962; 8:37–55.

[64] Immelmann K. Beiträge zu einer vergleichenden Biologie australischer Prachtfinken (Spermestidae). Zool Jb Sys Bd 1962:1–96.

[65] Dunn AM, Zann RA. Undirected song in wild zebra finch flocks: Contexts and effects of mate removal. Ethology 1996; 102:529–39. doi:10.1111/j.1439-0310.1996.tb01145.x.

[66] Miller JE, Hilliard AT, White SA. Song practice promotes acute vocal variability at a key stage of sensorimotor learning. PLoS ONE 2010; 5:e8592. doi:10.1371/journal.pone.0008592.

[67] Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics 2008; 9:559. doi:10.1186/1471-2105-9-559.

[68] Tchernichovski O, Nottebohm F, Ho C, Pesaran B, Mitra P. A procedure for an automated measurement of song similarity. Anim Behav 2000; 59:1167–76. doi:10.1006/anbe.1999.1416.

[69] Burkett ZD, Day NF, Peñagarikano O, Geschwind DH, White SA. VoICE: A semi-automated pipeline for standardizing vocal analysis across models. Sci Rep 2015; 5:10237. doi:10.1038/srep10237.

[70] Thompson CK, Schwabe F, Schoof A, Mendoza E, Gampe J, Rochefort C, Scharff C. Young and intense: FoxP2 immunoreactivity in Area X varies with age, song stereotypy, and singing in male zebra finches. Front Neural Circuits 2013; 7:24. doi:10.3389/fncir.2013.00024.

[71] Schneider CA, Rasband WS, Eliceiri KW. NIH Image to ImageJ: 25 years of image analysis. Nat Methods 2012; 9:671–5.

[72] Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M,

Gingeras TR. STAR: ultrafast universal RNA-seq aligner. Bioinformatics 2013; 29:15–21. doi:10.1093/bioinformatics/bts635.

[73] Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics 2014; 30:923–30. doi:10.1093/bioinformatics/btt656.

[74] Liao Y, Smyth GK, Shi W. The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. Nucleic Acids Res 2013; 41:e108–8. doi:10.1093/nar/gkt214.

[75] Ben-Ari Fuchs S, Lieder I, Stelzer G, Mazor Y, Buzhor E, Kaplan S, Bogoch Y, Plaschkes I, Shitrit A, Rappaport N, Kohn A, Edgar R, Shenhav L, Safran M, Lancet D, Guan-Golan Y, Warshawsky D, Shtrichman R. GeneAnalytics: An integrative gene set analysis tool for next generation sequencing, RNAseq and microarray data. Omics 2016; 20:139–51. doi:10.1089/omi.2015.0168.

[76] Nelson CS, Fuller CK, Fordyce PM, Greninger AL, Li H, DeRisi JL. Microfluidic affinity and ChIP-seq analyses converge on a conserved FOXP2-binding motif in chimp and human, which enables the detection of evolutionarily novel targets. Nucleic Acids Res 2013; 41:5991–6004. doi:10.1093/nar/gkt259.

[77] Mathelier A, Fornes O, Arenillas DJ, Chen C-Y, Denay G, Lee J, Shi W, Shyr C, Tan G, Worsley-Hunt R, Zhang AW, Parcy F, Lenhard B, Sandelin A, Wasserman WW. JASPAR 2016: a major expansion and update of the open-access database of transcription factor binding profiles. Nucleic Acids Res 2016; 44:D110–5. doi:10.1093/nar/gkv1176.

[78] Olias P, Adam I, Meyer A, Scharff C, Gruber AD. Reference genes for quantitative gene expression studies in multiple avian species. PLoS ONE 2014; 9:e99678. doi:10.1371/journal.pone.0099678.

[79] Feenders G, Liedvogel M, Rivas M, Zapka M, Horita H, Hara E, Wada K, Mouritsen H, Jarvis ED. Molecular mapping of movement-associated areas in the avian brain: A motor theory for vocal learning origin. PLoS ONE 2008; 3:e1768–27. doi:10.1371/journal.pone.0001768.

[80] Barrett HC. A hierarchical model of the evolution of human brain specializations. Proc Natl Acad Sci USA 2012; 109 Suppl 1:10733–40. doi:10.1073/pnas.1201898109.

[81] Oakley TH, Rivera AS. Genomics and the evolutionary origins of nervous system complexity. Curr Opin Genet Dev 2008; 18:479–92. doi:10.1016/j.gde.2008.12.002.

[82] Langfelder P, Luo R, Oldham MC, Horvath S. Is my network module preserved and reproducible? PLoS Comput Biol 2011; 7:e1001057. doi:10.1371/journal.pcbi.1001057.

[83] Enard W, Gehre S, Hammerschmidt K, Hölter SM, Blass T, Somel M, Brückner MK, Schreiweis C, Winter C, Sohr R, Becker L, Wiebe V, Nickel B, Giger T, Müller U, Groszer M, Adler T, Aguilar A, Bolle I, Calzada-Wack J, Dalke C, Ehrhardt N, Favor J, Fuchs H, Gailus-Durner V, Hans W, Hölzlwimmer G, Javaheri A, Kalaydjiev S, Kallnik M, Kling E, Kunder S, Mossbrugger I, Naton B, Racz I, Rathkolb B, Rozman J, Schrewe A, Busch DH, Graw J, Ivandic B, Klingenspor M, Klopstock T, Ollert M, Quintanilla-Martinez L, Schulz H, Wolf E, Wurst W, Zimmer A, Fisher SE, Morgenstern R, Arendt T, de Angelis MH, Fischer J, Schwarz J, Pääbo S. A humanized version of Foxp2 affects cortico-basal ganglia circuits in mice. Cell 2009; 137:961–71. doi:10.1016/j.cell.2009.03.041.

[84] Reimold AM, Grusby MJ, Kosaras B, Fries JWU, Mori R, Maniwa S, Clauss IM, Collins T, Sidman RL, Glimcher MJ, Glimcher LH. Chondrodysplasia and neurological abnormalities in ATF-2-deficient mice. Nature 1996; 379:262–5. doi:10.1038/379262a0.

[85] Bruhat A, Chérasse Y, Maurin A-C, Breitwieser W, Parry L, Deval C, Jones N, Jousse C, Fafournoux P. ATF2 is required for amino acid-regulated transcription by orchestrating specific histone acetylation. Nucleic Acids Res 2007; 35:1312–21. doi:10.1093/nar/gkm038.

[86] Kawasaki H, Schiltz L, Chiu R, Itakura K, Taira K, Nakatani Y, Yokoyama KK. ATF-2 has intrinsic histone acetyltransferase activity which is modulated by phosphorylation. Nature 2000; 405:195–200. doi:10.1038/35012097.

[87] Bousiges O, Neidl R, Majchrzak M, Muller M-A, Barbelivien A, Pereira de Vasconcelos A, Schneider A, Loeffler J-P, Cassel J-C, Boutillier A-L. Detection of histone acetylation levels in the dorsal hippocampus reveals early tagging on specific residues of H2B and H4 histones in response to learning. PLoS ONE 2013; 8:e57816. doi:10.1371/journal.pone.0057816.

[88] Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, Simonovic M, Roth A, Santos A, Tsafou KP, Kuhn M, Bork P, Jensen LJ, Mering von C. STRING v10: protein-protein interaction networks, integrated over the tree of life. Nucleic Acids Res 2015; 43:D447–52. doi:10.1093/nar/gku1003.

[89] Chen Q, Heston JB, Burkett ZD, White SA. Expression analysis of the speech-related genes FoxP1 and FoxP2 and their relation to singing behavior in two songbird species. J Exp Biol 2013; 216:3682–92. doi:10.1242/jeb.085886.

[90] Winograd C, Ceman S. Exploring the zebra finch Taeniopygia guttata as a novel animal model for the speech-language deficit of fragile X syndrome. Results Probl Cell Differ 2012; 54:181–97. doi:10.1007/978-3-642-21649-7_10.

[91] Winograd C, Clayton D, Ceman S. Expression of fragile X mental retardation protein within the vocal control system of developing and adult male zebra finches. Neuroscience 2008; 157:132–42. doi:10.1016/j.neuroscience.2008.09.005.

[92] Ascano M, Mukherjee N, Bandaru P, Miller JB, Nusbaum JD, Corcoran DL, Langlois C, Munschauer M, Dewell S, Hafner M, Williams Z, Ohler U, Tuschl T. FMRP targets distinct mRNA sequence elements to regulate protein expression. Nature 2012; 492:382–6. doi:10.1038/nature11737.

[93] Enslen H, Raingeaud J, Davis RJ. Selective Activation of p38 Mitogen-activated Protein (MAP) Kinase Isoforms by the MAP Kinase Kinases MKK3 and MKK6. J Biol Chem 1998; 273:1741–8. doi:10.1074/jbc.273.3.1741.

[94] Stein B, Yang MX, Young DB, Janknecht R, Hunter T, Murray BW, Barbosa MS. p38-2, a novel mitogen-activated protein kinase with distinct properties. J Biol Chem 1997; 272:19509–17.

[95] Dong J, Horvath S. Understanding network concepts in modules. BMC Syst Biol 2007; 1:24. doi:10.1186/1752-0509-1-24.

[96] Zhao W, Langfelder P, Fuller T, Dong J, Li A, Hovarth S. Weighted gene coexpression network analysis: state of the art. J Biopharm Stat 2010; 20:281–300. doi:10.1080/10543400903572753.

[97] Yip AM, Horvath S. Gene network interconnectedness and the generalized topological overlap measure. BMC Bioinformatics 2007; 8:22. doi:10.1186/1471-2105-8-22.

[98] Horvath S. Weighted Network Analysis. New York: Springer Science & Business Media; 2011.

[99] Neve RL, Neve KA, Nestler EJ, Carlezon WA. Use of herpes virus amplicon vectors to study brain disorders. BioTechniques 2005; 39:381–91.

[100] Jacobs EC, Arnold AP, Campagnoni AT. Developmental regulation of the distribution of aromatase- and estrogen-receptor- mRNA-expressing cells in the zebra finch Brain. Dev Neurosci 2000; 21:453–72. doi:10.1159/000017413.

[101] Ye J, Coulouris G, Zaretskaya I, Cutcutache I, Rozen S, Madden TL. Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. BMC Bioinformatics 2012; 13:134. doi:10.1186/1471-2105-13-134.

[102] Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2−ΔΔCT method. Methods 2001; 25:402–8. doi:10.1006/meth.2001.1262.

[103] Miller JE, Hafzalla GW, Burkett ZD, Fox CM, White SA. Reduced vocal variability in a zebra finch model of dopamine depletion: implications for Parkinson disease. Physiol Rep 2015; 3:e12599–9. doi:10.14814/phy2.12599.

[104] Teramitsu I, Poopatanapong A, Torrisi S, White SA. Striatal FoxP2 is actively regulated during songbird sensorimotor learning. PLoS ONE 2010; 5:e8548. doi:10.1371/journal.pone.0008548.

[105] Brainard MS, Doupe AJ. Translating birdsong: songbirds as a model for basic and applied medical research. Annu Rev Neurosci 2013; 36:489–517. doi:10.1146/annurev-neuro-060909-152826.

[106] Mahrt EJ, Perkel DJ, Tong L, Rubel EW, Portfors CV. Engineered deafness reveals that mouse courtship vocalizations do not require auditory experience. J Neurosci 2013; 33:5573–83.

[107] Day NF, Fraley ER. Insights from a nonvocal learner on social communication. J Neurosci 2013; 33:12553–4. doi:10.1523/jneurosci.2258-13.2013.

[108] Brenowitz EA, Margoliash D, Nordeen KW. An introduction to birdsong and the avian song system. J Neurobiol 1997; 33:495–500.

[109] Wu W, Thompson JA, Bertram R, Johnson F. A statistical method for quantifying songbird phonology and syntax. J Neurosci Methods 2008; 174:147–54. doi:10.1016/j.jneumeth.2008.06.033.

[110] Daou A, Johnson F, Wu W, Bertram R. A computational tool for automated large-scale analysis and measurement of bird-song syntax. J Neurosci Methods 2012; 210:147–60. doi:10.1016/j.jneumeth.2012.07.020.

[111] Tachibana RO, Oosugi N, Okanoya K. Semi-automatic classification of birdsong elements using a linear support vector machine. PLoS ONE 2014; 9:e92584. doi:10.1371/journal.pone.0092584.

[112] Mandelblat-Cerf Y, Fee MS. An automated procedure for evaluating song imitation. PLoS ONE 2014; 9:e96484. doi:10.1371/journal.pone.0096484.

[113] Portfors CV. Types and functions of ultrasonic vocalizations in laboratory rats and mice. J Am Assoc Lab Anim Sci 2007; 46:28–34.

[114] Scattoni ML, Gandhy SU, Ricceri L, Crawley JN. Unusual repertoire of vocalizations in the BTBR T+tf/J mouse model of autism. PLoS ONE 2008; 3:e3067. doi:10.1371/journal.pone.0003067.

[115] Grimsley JMS, Gadziola MA, Wenstrup JJ. Automated classification of mouse pup isolation syllables: from cluster analysis to an Excel-based "mouse pup syllable classification calculator". Front Behav Neurosci 2012; 6:89. doi:10.3389/fnbeh.2012.00089.

[116] Peñagarikano O, Abrahams BS, Herman EI, Winden KD, Gdalyahu A, Dong H, Sonnenblick LI, Gruver R, Almajano J, Bragin A, Golshani P, Trachtenberg JT, Peles E, Geschwind DH. Absence of CNTNAP2 leads to epilepsy, neuronal migration abnormalities, and core autism-related deficits. 2011; 147:235–46. doi:10.1016/j.cell.2011.08.040.

[117] Konishi M. The role of auditory feedback in the vocal behavior of the domestic fowl. Z Tierpsychol 1963; 20:349–67.

[118] Nordeen KW, Nordeen EJ. Deafening-induced vocal deterioration in adult songbirds is reversed by disrupting a basal ganglia-forebrain circuit. J Neurosci 2010; 30:7392–400. doi:10.1523/jneurosci.6181-09.2010.

[119] Langfelder P, Horvath S. Fast R functions for robust correlations and hierarchical clustering. J Stat Softw 2012; 46.

[120] Sossinka R, Böhner J. Song types in the zebra finch Poephila guttata castanotis. Z Tierpsychol 1980; 53:123–32.

[121] Horita H, Wada K, Jarvis ED. Early onset of deafening-induced song deterioration and differential requirements of the pallial-basal ganglia vocal pathway. Eur J Neurosci 2008; 28:2519–32. doi:10.1111/j.1460-9568.2008.06535.x.

[122] Nordeen KW, Nordeen EJ. Auditory feedback is necessary for the maintenance of stereotyped song in adult zebra finches. Behav Neural Biol 1992; 57:58–66. doi:10.1016/0163-1047(92)90757-U.

[123] Brainard MS, Doupe AJ. Interruption of a basal ganglia-forebrain circuit prevents plasticity of learned vocalizations. Nature 2000; 404:762–6. doi:10.1038/35008083.

[124] Thompson JA, Wu W, Bertram R, Johnson F. Auditory-dependent vocal recovery in adult male zebra finches is facilitated by lesion of a forebrain pathway that includes the basal ganglia. J Neurosci 2007; 27:12308–20. doi:10.1523/jneurosci.2853-07.2007.

[125] Crawley JN. What's wrong with my mouse?: Behavioral Phenotyping of Transgenic and Knockout Mice. 2nd ed. Hoboken: John Wiley & Sons; 2007.

[126] Silverman JL, Yang M, Lord C, Crawley JN. Behavioural phenotyping assays for mouse models of autism. Nat Rev Neurosci 2010; 11:490–502. doi:10.1038/nrn2851.

[127] Geschwind DH, Levitt P. Autism spectrum disorders: developmental disconnection syndromes. Curr Opin Neurobiol 2007; 17:103–11. doi:10.1016/j.conb.2007.01.009.

[128] Scott BB, Velho TA, Sim S, Lois C. Applications of avian transgenesis. Ilar J 2010; 51:353–61. doi:10.1093/ilar.51.4.353.

[129] Gkogkas CG, Khoutorsky A, Ran I, Rampakakis E, Nevarko T, Weatherill DB, Vasuta C, Yee S, Truitt M, Dallaire P, Major F, Lasko P, Ruggero D, Nader K, Lacaille J-C, Sonenberg N. Autism-related deficits via dysregulated eIF4E-dependent translational control. Nature 2013; 493:371–7. doi:10.1038/nature11628.

[130] Santini E, Huynh TN, MacAskill AF, Carter AG, Pierre P, Ruggero D, Kaphzan H, Klann E. Exaggerated translation causes synaptic and behavioural aberrations associated with autism. Nature 2013; 493:411–5. doi:10.1038/nature11782.

[131] Lai JKY, Sobala-Drozdowski M, Zhou L, Doering LC, Faure PA, Foster JA. Temporal and spectral differences in the ultrasonic vocalizations of fragile X knock out mice during postnatal development. Behav Brain Res 2014; 259:119–30. doi:10.1016/j.bbr.2013.10.049.

[132] Nakatani J, Tamada K, Hatanaka F, Ise S, Ohta H, Inoue K, Tomonaga S, Watanabe Y, Chung YJ, Banerjee R, Iwamoto K, Kato T, Okazawa M, Yamauchi K, Tanda K, Takao K, Miyakawa T, Bradley A, Takumi T. Abnormal behavior in a chromosome-engineered mouse model for human 15q11-13 duplication seen in autism. Cell 2009; 137:1235–46. doi:10.1016/j.cell.2009.04.024.

[133] De Rubeis S, Bagni C. Regulation of molecular pathways in the Fragile X Syndrome: insights into autism spectrum disorders. J Neurodev Disord 2011; 3:257.

doi:10.1007/s11689-011-9087-2.

[134] Moon J, Beaudin AE, Verosky S, Driscoll LL, Weiskopf M, Levitsky DA, Crnic LS, Strupp BJ. Attentional dysfunction, impulsivity, and resistance to change in a mouse model of fragile X syndrome. Behav Neurosci 2006; 120:1367–79. doi:10.1037/0735-7044.120.6.1367.

[135] MI Z, GL V, GKh M. Differences in behavior of impulsive and self-controlled rats in the open-field and light-dark tests. Zh Vyssh Nerv Deiat Im I P Pavlova 2011; 61:340–50.

[136] Phillips RG, LeDoux JE. Differential contribution of amygdala and hippocampus to cued and contextual fear conditioning. Behav Neurosci 1992; 106:274–85. doi:10.1037/0735-7044.106.2.274.

[137] Kim JJ, Fanselow MS. Modality-specific retrograde amnesia of fear. Science 1992; 256:675–7.

[138] Kedia S, Chattarji S. Marble burying as a test of the delayed anxiogenic effects of acute immobilisation stress in mice. J Neurosci Methods 2014; 233:150–4. doi:10.1016/j.jneumeth.2014.06.012.

[139] Njung'e K, Handley SL. Effects of 5-HT uptake inhibitors, agonists and antagonists on the burying of harmless objects by mice; a putative test for anxiolytic agents. Br J Pharmacol 1991; 104:105–12. doi:10.1111/j.1476-5381.1991.tb12392.x.

[140] Kaelbling, L. P., Littman ML, Moore AW. Reinforcement learning: A survey. J Artif Intell Res 1996; 4:237–85. doi:10.1613/jair.301.

[141] Wu HG, Miyamoto YR, Gonzalez Castro LN, Ölveczky BP, Smith MA. Temporal structure of motor variability is dynamically regulated and predicts motor learning ability. Nat Neurosci 2014; 17:312–21. doi:10.1038/nn.3616.

[142] Derégnaucourt S, Mitra PP, Fehér O, Pytte C, Tchernichovski O. How sleep affects the developmental learning of bird song. Nature 2005; 433:710–6. doi:10.1038/nature03275.

[143] Tumer EC, Brainard MS. Performance variability enables adaptive plasticity of "crystallized" adult birdsong. Nature 2007; 450:1240–4. doi:10.1038/nature06390.

[144] Sober SJ, Brainard MS. Vocal learning is constrained by the statistics of sensorimotor

experience. Proc Natl Acad Sci USA 2012; 109:21099–103. doi:10.1073/pnas.1213622109.

[145] Wu HG, Miyamoto YR, Castro LNG, Ölveczky BP, Smith MA. Temporal structure of motor variability is dynamically regulated and predicts motor learning ability. Nat Neurosci 2014; 17:312–21. doi:10.1038/nn.3616.

[146] Haesler S, Wada K, Nshdejan A, Morrisey EE, Lints T, Jarvis ED, Scharff C. FoxP2 expression in avian vocal learners and non-learners. J Neurosci 2004; 24:3164–75. doi:10.1523/jneurosci.4369-03.2004.

[147] Wang B, Lin D, Li C, Tucker P. Multiple domains define the expression and regulatory properties of Foxp1 forkhead transcriptional repressors. J Biol Chem 2003; 278:24259–68. doi:10.1074/jbc.M207174200.

[148] Li S, Weidenfeld J, Morrisey EE. Transcriptional and DNA binding activity of the Foxp1/2/4 family Is modulated by heterotypic and homotypic protein interactions. Mol Cell Biol 2004; 24:809–22. doi:10.1128/mcb.24.2.809-822.2004.

[149] Murugan M, Harward S, Scharff C, Mooney R. Diminished FoxP2 levels affect dopaminergic modulation of corticostriatal signaling important to song variability. Neuron 2013; 80:1464–76. doi:10.1016/j.neuron.2013.09.021.

[150] Onn S-P, Fienberg AA, Grace AA. Dopamine modulation of membrane excitability in striatal spiny neurons is altered in DARPP-32 knockout mice. J Pharmacol Exp Ther 2003; 306:870–9. doi:10.1124/jpet.103.050062.

[151] Leblois A, Perkel DJ. Striatal dopamine modulates song spectral but not temporal features through D1 receptors. Eur J Neurosci 2012; 35:1771–81. doi:10.1111/j.1460-9568.2012.08095.x.

[152] Stroud JC, Wu Y, Bates DL, Han A, Nowick K, Pääbo S, Tong H, Chen L. Structure of the forkhead domain of FOXP2 bound to DNA. Structure 2006; 14:159–66. doi:10.1016/j.str.2005.10.005.

[153] Adegbola AA, Cox GF, Bradshaw EM, Hafler DA, Gimelbrant A, Chess A. Monoallelic expression of the human FOXP2 speech gene. Proc Natl Acad Sci USA 2015; 112:6848–54. doi:10.1073/pnas.1411270111.

[154] Shah S, Lubeck E, Zhou W, Cai L. In situ transcription profiling of single cells reveals spatial organization of cells in the mouse hippocampus. Neuron 2016; 92:342–57.

doi:10.1016/j.neuron.2016.10.001.

[155] Alcock KJ, Passingham RE, Watkins KE, Vargha-Khadem F. Oral dyspraxia in inherited speech and language impairment and acquired dysphasia. Brain Lang 2000; 75:17–33. doi:10.1006/brln.2000.2322.

[156] Watkins KE, Dronkers NF, Vargha-Khadem F. Behavioural analysis of an inherited speech and language disorder: comparison with acquired aphasia. Brain 2002; 125:452–64. doi:10.1093/brain/awf058.

[157] Shu W, Yang H, Zhang L, Lu MM, Morrisey EE. Characterization of a new subfamily of winged-helix/forkhead (Fox) genes that are expressed in the lung and act as transcriptional repressors. J Biol Chem 2001; 276:27488–97. doi:10.1074/jbc.M100636200.

[158] Carr CW, Moreno-De-Luca D, Parker C, Zimmerman HH, Ledbetter N, Martin CL, Dobyns WB, Abdul-Rahman OA. Chiari I malformation, delayed gross motor skills, severe speech delay, and epileptiform discharges in a child with FOXP1 haploinsufficiency. Eur J Hum Genet 2010; 18:1216–20. doi:10.1038/ejhg.2010.96.

[159] Hamdan FF, Daoud H, Rochefort D, Piton A, Gauthier J, Langlois M, Foomani G, Dobrzeniecka S, Krebs M-O, Joober R, Lafrenière RG, Lacaille J-C, Mottron L, Drapeau P, Beauchamp MH, Phillips MS, Fombonne E, Rouleau GA, Michaud JL. De novo mutations in FOXP1 in cases with intellectual disability, autism, and language impairment. Am J Hum Genet 2010; 87:671–8. doi:10.1016/j.ajhg.2010.09.017.

[160] Horn D, Kapeller J, Rivera-Brugués N, Moog U, Lorenz-Depiereux B, Eck S, Hempel M, Wagenstaller J, Gawthrope A, Monaco AP, Bonin M, Riess O, Wohlleber E, Illig T, Bezzina CR, Franke A, Spranger S, Villavicencio-Lorini P, Seifert W, Rosenfeld J, Klopocki E, Rappold GA, Strom TM. Identification of FOXP1 deletions in three unrelated patients with mental retardation and significant speech and language deficits. Hum Mutat 2010; 31:E1851–60. doi:10.1002/humu.21362.

[161] Pariani MJ, Spencer A, Graham JM Jr., Rimoin DL. A 785kb deletion of 3p14.1p13, including the FOXP1 gene, associated with speech delay, contractures, hypertonia and blepharophimosis. Eur J Med Genet 2009; 52:123–7. doi:10.1016/j.ejmg.2009.03.012.

[162] Kikusui T, Nakanishi K, Nakagawa R, Nagasawa M, Mogi K, Okanoya K. Cross fostering experiments suggest that mice songs are innate. PLoS ONE 2011; 6:e17721. doi:10.1371/journal.pone.0017721.

[163] Price PH. Developmental determinants of structure in zebra finch song. J Comp Physiol

Psychol 1979; 93:260–77. doi:10.1037/h0077553.

[164] Scharff C, Nottebohm F. A comparative study of the behavioral deficits following lesions of various parts of the zebra finch song system: implications for vocal learning. J Neurosci 1991; 11:2896–913.

[165] Brainard MS, Doupe AJ. Auditory feedback in learning and maintenance of vocal behaviour. Nat Rev Neurosci 2000; 1:31–40. doi:10.1038/35036205.

[166] Nottebohm F, Stokes TM, Leonard CM. Central control of song in the canary, Serinus canarius. J Comp Neurol 1976; 165:457–86. doi:10.1002/cne.901650405.

[167] Nottebohm F. The neural basis of birdsong. PLoS Biol 2005; 3:e164. doi:10.1371/journal.pbio.0030164.

[168] Gale SD, Perkel DJ. Anatomy of a songbird basal ganglia circuit essential for vocal learning and plasticity. J Chem Neuroanat 2010; 39:124–31. doi:10.1016/j.jchemneu.2009.07.003.

[169] Okanoya K. The Bengalese finch: A window on the behavioral neurobiology of birdsong syntax. Ann N Y Acad Sci 2004; 1016:724–35. doi:10.1196/annals.1298.026.

[170] Okanoya K, Yamaguchi A. Adult Bengalese finches (Lonchura striata var. domestica) require real-time auditory feedback to produce normal song syntax. J Neurobiol 1997; 33:343–56.

[171] Woolley SMN, Rubel EW. Bengalese finches Lonchura striata domestica depend upon auditory feedback for the maintenance of adult song. J Neurosci 1997; 17:6380–90. doi:10.1016/0003-3472(69)90157-2.

[172] Miller JE, Spiteri E, Condro MC, Dosumu-Johnson RT, Geschwind DH, White SA. Birdsong decreases protein levels of FoxP2, a molecule required for human speech. J Neurophysiol 2008; 100:2015–25. doi:10.1152/jn.90415.2008.

[173] Jürgens U. The neural control of vocalization in mammals: a review. J Voice 2009; 23:1–10. doi:10.1016/j.jvoice.2007.07.005.

[174] Rousso DL, Gaber ZB, Wellik D, Morrisey EE, Novitch BG. Coordinated actions of the forkhead protein Foxp1 and Hox proteins in the columnar organization of spinal motor neurons. Neuron 2008; 59:226–40. doi:10.1016/j.neuron.2008.06.025.

[175] Wild JM, Farabaugh SM. Organization of afferent and efferent projections of the nucleus basalis prosencephali in a passerine, Taeniopygia guttata. J Comp Neurol 1996; 365:306–28. doi:10.1002/(SICI)1096-9861(19960205)365:2<306::AID-CNE8>3.0.CO;2-9.

[176] Bottjer SW, Miesner EA, Arnold AP. Forebrain lesions disrupt development but not maintenance of song in passerine birds. Science 1984; 224:901–3.

[177] Whitney O, Johnson F. Motor-induced transcription but sensory-regulated translation of ZENK in socially interactive songbirds. J Neurobiol 2005; 65:251–9. doi:10.1002/neu.20187.

[178] Carlsson P, Mahlapuu M. Forkhead transcription factors: Key players in development and metabolism. Dev Biol 2002; 250:1–23. doi:10.1006/dbio.2002.0780.

[179] Clovis YM, Enard W, Marinaro F, Huttner WB, De Pietri Tonelli D. Convergent repression of Foxp2 3′UTR by miR-9 and miR-132 in embryonic mouse neocortex: implications for radial migration of neurons. Development 2012; 139:3332–42. doi:10.1242/dev.078063.

[180] Brainard MS, Doupe AJ. Postlearning consolidation of birdsong: Stabilizing effects of age and anterior forebrain lesions. J Neurosci 2001; 21:2501–17.

[181] Cooper BG, Méndez JM, Saar S, Whetstone AG, Meyers R, Goller F. Age-related changes in the Bengalese finch song motor program. Neurobiol Aging 2012; 33:564–8. doi:10.1016/j.neurobiolaging.2010.04.014.

[182] Nixdorf-Bergweiler BE, Bischof H-J. A stereotaxic atlas of the brain of the zebra finch, Taeniopygia guttata, with special emphasis on telencephalic visual and song system nuclei in transverse and sagittal sections. Bethesda: National Library of Medicine, National Center for Biotechnology Information; 2007.

[183] Bottjer SW. The distribution of tyrosine hydroxylase immunoreactivity in the brains of male and female zebra finches. Dev Neurobiol 1993; 24:51–69. doi:10.1002/neu.480240105.

[184] Lewis JW, Ryan SM, Arnold AP, Butcher LL. Evidence for a catecholaminergic projection to area X in the zebra finch. J Comp Neurol 1981; 196:347–54. doi:10.1002/cne.901960212.

[185] Person AL, Gale SD, Farries MA, Perkel DJ. Organization of the songbird basal ganglia,

including area X. J Comp Neurol 2008; 508:840–66. doi:10.1002/cne.21699.

[186] Gale SD, Perkel DJ. Anatomy of a songbird basal ganglia circuit essential for vocal learning and plasticity. J Chem Neuroanat 2010; 39:124–31. doi:10.1016/j.jchemneu.2009.07.003.

[187] Shiflett MW, Balleine BW. Molecular substrates of action control in cortico-striatal circuits. Prog Neurobiol 2011; 95:1–13. doi:10.1016/j.pneurobio.2011.05.007.

[188] Simonyan K, Horwitz B, Jarvis ED. Dopamine regulation of human speech and bird song: A critical review. Brain Lang 2012; 122:142–50. doi:10.1016/j.bandl.2011.12.009.

[189] Goldberg JH, Fee MS. Singing-related neural activity distinguishes four classes of putative striatal neurons in the songbird basal ganglia. J Neurophysiol 2010; 103:2002–14. doi:10.1152/jn.01038.2009.

[190] Heimovics SA, Riters LV. Evidence that dopamine within motivation and song control brain regions regulates birdsong context-dependently. Physiol Behav 2008; 95:258–66. doi:10.1016/j.physbeh.2008.06.009.

[191] Sasaki A, Sotnikova TD, Gainetdinov RR, Jarvis ED. Social context-dependent singing-regulated dopamine. J Neurosci 2006; 26:9010–4. doi:10.1523/jneurosci.1335-06.2006.

[192] Hara E, Kubikova L, Hessler NA, Jarvis ED. Role of the midbrain dopaminergic system in modulation of vocal brain activation by social context. Eur J Neurosci 2007; 25:3406–16. doi:10.1111/j.1460-9568.2007.05600.x.

[193] Yanagihara S, Hessler NA. Modulation of singing-related activity in the songbird ventral tegmental area by social context. Eur J Neurosci 2006; 24:3619–27. doi:10.1111/j.1460-9568.2006.05228.x.

[194] Grant LM, Barnett DG, Doll EJ, Leverson G, Ciucci M. Relationships among rat ultrasonic vocalizations, behavioral measures of striatal dopamine loss, and striatal tyrosine hydroxylase immunoreactivity at acute and chronic time points following unilateral 6-hydroxydopamine-induced dopamine depletion. Behav Brain Res 2015; 291:361–71. doi:10.1016/j.bbr.2015.05.042.

[195] Leblois A, Wendel BJ, Perkel DJ. Striatal dopamine modulates basal ganglia output and regulates social context-dependent behavioral variability through D1 receptors. J Neurosci 2010; 30:5730–43. doi:10.1523/jneurosci.5974-09.2010.

[196] Ciucci MR, Schaser AJ, Russell JA. Exercise-induced rescue of tongue function without striatal dopamine sparing in a rat neurotoxin model of Parkinson disease. Behav Brain Res 2013; 252:239–45. doi:10.1016/j.bbr.2013.06.004.

[197] Gale SD, Perkel DJ. Properties of dopamine release and uptake in the songbird basal ganglia. J Neurophysiol 2005; 93:1871–9. doi:10.1152/jn.01053.2004.

[198] Kirik D, Rosenblad C, Bjorklund A. Characterization of behavioral and neurodegenerative changes following partial lesions of the nigrostriatal dopamine system induced by intrastriatal 6-hydroxydopamine in the rat. Exp Neurol 1998; 152:259–77. doi:10.1006/exnr.1998.6848.

[199] Morales I, Sanchez A, Rodriguez-Sabate C, Rodriguez M. The degeneration of dopaminergic synapses in Parkinson's disease: A selective animal model. Behav Brain Res 2015; 289:19–28. doi:10.1016/j.bbr.2015.04.019.

[200] Przedborski S, Levivier M, Jiang H, Ferreira M, Jackson-Lewis V, Donaldson D, Togasaki DM. Dose-dependent lesions of the dopaminergic nigrostriatal pathway induced by intrastriatal injection of 6-hydroxydopamine. Neuroscience 1995; 67:631–47.

[201] Ciucci MR, Ma ST, Fox C, Kane JR, Ramig LO, Schallert T. Qualitative changes in ultrasonic vocalization in rats after unilateral dopamine depletion or haloperidol: a preliminary study. Behav Brain Res 2007; 182:284–9. doi:10.1016/j.bbr.2007.02.020.

[202] Ciucci MR, Ahrens AM, Ma ST, Kane JR, Windham EB, Woodlee MT, Schallert T. Reduction of dopamine synaptic activity: degradation of 50-kHz ultrasonic vocalization in rats. Behav Neurosci 2009; 123:328–36. doi:10.1037/a0014593.

[203] Sapir S. Multiple factors are involved in the dysarthria associated with Parkinson's disease: a review with implications for clinical practice and research. J Speech Lang Hear Res 2014; 57:1330–43. doi:10.1044/2014_JSLHR-S-13-0039.

[204] Harel B, Cannizzaro M, Snyder PJ. Variability in fundamental frequency during speech in prodromal and incipient Parkinson's disease: a longitudinal case study. Brain Cogn 2004; 56:24–9. doi:10.1016/j.bandc.2004.05.002.

[205] Ho AK, Bradshaw JL, Iansek R, Alfredson R. Speech volume regulation in Parkinson's disease: effects of implicit cues and explicit instructions. Neuropsychologia 1999; 37:1453–60.

[206] Qin ZH, Chen JF, Weiss B. Lesions of mouse striatum induced by 6-hydroxydopamine differentially alter the density, rate of synthesis, and level of gene expression of D1 and D2 dopamine receptors. J Neurochem 1994; 62:411–20.

[207] Kubikova L, Wada K, Jarvis ED. Dopamine receptors in a songbird brain. J Comp Neurol 2010; 518:741–69. doi:10.1002/cne.22255.

[208] Calabresi P, Picconi B, Tozzi A, Ghiglieri V, Di Filippo M. Direct and indirect pathways of basal ganglia: a critical reappraisal. Nat Neurosci 2014; 17:1022–30. doi:10.1038/nn.3743.

[209] Kupchik YM, Brown RM, Heinsbroek JA, Lobo MK, Schwartz DJ, Kalivas PW. Coding the direct/indirect pathways by D1 and D2 receptors is not valid for accumbens projections. Nat Neurosci 2015; 18:1230–2. doi:10.1038/nn.4068.

[210] Cornil CA, Castelino CB, Ball GF. Dopamine binds to α2-adrenergic receptors in the song control system of zebra finches (Taeniopygia guttata). J Chem Neuroanat 2008; 35:202–15. doi:10.1016/j.jchemneu.2007.10.004.

[211] Castelino CB, Schmidt MF. What birdsong can teach us about the central noradrenergic system. J Chem Neuroanat 2010; 39:96–111. doi:10.1016/j.jchemneu.2009.08.003.

[212] Braak H, Del Tredici K, Rüb U, de Vos RAI, Jansen Steur ENH, Braak E. Staging of brain pathology related to sporadic Parkinson's disease. Neurobiol Aging 2003; 24:197–211.

[213] Hilliard AT, Miller JE, Horvath S, White SA. Distinct neurogenomic states in basal ganglia subregions relate differently to singing behavior in songbirds. PLoS Comput Biol 2012; 8:e1002773. doi:10.1371/journal.pcbi.1002773.

[214] D'Arcangelo G, Miao GG, Chen SC, Soares HD, Morgan JI, Curran T. A protein related to extracellular matrix proteins deleted in the mouse mutant reeler. Nature 1995; 374:719–23. doi:10.1038/374719a0.

[215] Weeber EJ, Beffert U, Jones C, Christian JM, Forster E, Sweatt JD, Herz J. Reelin and ApoE receptors cooperate to enhance hippocampal synaptic plasticity and learning. J Biol Chem 2002; 277:39944–52. doi:10.1074/jbc.M205147200.

[216] Niu S, Renfro A, Quattrocchi CC, Sheldon M, D'Arcangelo G. Reelin promotes hippocampal dendrite development through the VLDLR/ApoER2-Dab1 pathway. Neuron 2004; 41:71–84.

[217] Niu S, Yabut O, D'Arcangelo G. The Reelin signaling pathway promotes dendritic spine development in hippocampal neurons. J Neurosci 2008; 28:10339–48. doi:10.1523/jneurosci.1917-08.2008.

[218] Rice DS, Curran T. Role of the reelin signaling pathway in central nervous system development. Annu Rev Neurosci 2001; 24:1005–39. doi:10.1146/annurev.neuro.24.1.1005.

[219] Herz J, Chen Y. Reelin, lipoprotein receptors and synaptic plasticity. Nat Rev Neurosci 2006; 7:850–9. doi:10.1038/nrn2009.

[220] Goffinet AM, So KF, Yamamoto M, Edwards M, Caviness VS. Architectonic and hodological organization of the cerebellum in reeler mutant mice. Brain Res 1984; 318:263–76.

[221] Caviness VS, Rakic P. Mechanisms of cortical development: a view from mutations in mice. Annu Rev Neurosci 1978; 1:297–326. doi:10.1146/annurev.ne.01.030178.001501.

[222] Falconer DS. Two new mutants, 'trembler' and “reeler,” with neurological actions in the house mouse (Mus musculus L.). J Genet 1951; 50:192–201.

[223] Tissir F, Goffinet AM. Reelin and brain development. Nat Rev Neurosci 2003; 4:496–505. doi:10.1038/nrn1113.

[224] D'Arcangelo G, Homayouni R, Keshvara L, Rice DS, Sheldon M, Curran T. Reelin is a ligand for lipoprotein receptors. Neuron 1999; 24:471–9.

[225] Howell BW, Hawkes R, Soriano P, Cooper JA. Neuronal position in the developing brain is regulated by mouse disabled-1. Nature 1997; 389:733–7. doi:10.1038/39607.

[226] Hiesberger T, Trommsdorff M, Howell BW, Goffinet A, Mumby MC, Cooper JA, Herz J. Direct binding of Reelin to VLDL Receptor and ApoE Receptor 2 induces tyrosine phosphorylation of Disabled-1 and modulates Tau phosphorylation. Neuron 1999; 24:481–9. doi:10.1016/S0896-6273(00)80861-2.

[227] Bock HH, Herz J. Reelin activates Src family tyrosine kinases in neurons. Current Biology 2003; 13:18–26.

[228] Persico AM, D'agruma L, Maiorano N, Totaro A, Militerni R, Bravaccio C, Wassink TH, Schneider C, Melmed R, Trillo S, Montecchi F, Palermo M, Pascucci T, Puglisi-

Allegra S, Reichelt KL, Conciatori M, Marino R, Quattrocchi CC, Baldi A, Zelante L, Gasparini P, Keller F. Reelin gene alleles and haplotypes as a factor predisposing to autistic disorder. Mol Psychiatry 2001; 6:150–9. doi:10.1038/sj.mp.4000850.

[229] Vorstman JAS, Morcus MEJ, Duijff SN, Klaassen PWJ, Heineman-de Boer JA, Beemer FA, Swaab H, Kahn RS, Van Engeland H. The 22q11.2 deletion in children. J Am Acad Child Adolesc Psychiatry 2006; 45:1104–13.

[230] Serajee FJ, Zhong H, Mahbubul Huq AHM. Association of Reelin gene polymorphisms with autism. Genomics 2006; 87:75–83.

[231] Skaar DA, Shao Y, Haines JL, Stenger JE, Jaworski J, Martin ER, DeLong GR, Moore JH, McCauley JL, Sutcliffe JS, Ashley-Koch AE, Cuccaro ML, Folstein SE, Gilbert JR, Pericak-Vance MA. Analysis of the RELN gene as a genetic risk factor for autism. Mol Psychiatry 2005; 10:563–71. doi:10.1038/sj.mp.4001614.

[232] Dutta S, Guhathakurta S, Sinha S, Chatterjee A, Ahmed S, Ghosh S, Gangopadhyay PK, Singh M, Usha R. Reelin gene polymorphisms in the Indian population: a possible paternal 5'UTR-CGG-repeat-allele effect on autism. Am J Med Genet B Neuropsychiatr Genet 2007; 144B:106–12. doi:10.1002/ajmg.b.30419.

[233] Kelemenova S, Schmidtova E, Ficek A, Celec P, Kubranska A, Ostatnikova D. Polymorphisms of candidate genes in Slovak autistic patients. Psychiatr Genet 2010; 20:137–9. doi:10.1097/YPG.0b013e32833a1eb3.

[234] Li H, Li Y, Shao J, Li R, Qin Y, Xie C, Zhao Z. The association analysis of RELN and GRM8 genes with autistic spectrum disorder in chinese han population. Am J Med Genet B Neuropsychiatr Genet 2008; 147B:194–200. doi:10.1002/ajmg.b.30584.

[235] Holt R, Barnby G, Maestrini E, Bacchelli E, Brocklebank D, Sousa I, Mulder EJ, Kantojärvi K, Järvelä I, Klauck SM, Poustka F, Bailey AJ, Monaco AP. Linkage and candidate gene studies of autism spectrum disorders in European populations. Eur J Hum Genet 2010; 18:1013–9. doi:10.1038/ejhg.2010.69.

[236] Ashley-Koch AE, Jaworski J, Ma DQ, Mei H, Ritchie MD, Skaar DA, Robert Delong G, Worley G, Abramson RK, Wright HH, Cuccaro ML, Gilbert JR, Martin ER, Pericak-Vance MA. Investigation of potential gene–gene interactions between apoe and reln contributing to autism risk. Psychiatr Genet 2007; 17:221–6. doi:10.1097/YPG.0b013e32809c2f75.

[237] Fatemi SH, Snow AV, Stary JM, Araghi-Niknam M, Reutiman TJ, Lee S, Brooks AI, Pearce DA. Reelin signaling is impaired in autism. Biol Psychiatry 2005; 57:777–87.

doi:10.1016/j.biopsych.2004.12.018.

[238] Zhubi A, Chen Y, Dong E, Cook EH, Guidotti A, Grayson DR. Increased binding of MeCP2 to the GAD1 and RELN promoters may be mediated by an enrichment of 5-hmC in autism spectrum disorder (ASD) cerebellum. Transl Psychiatry 2014; 4:e349. doi:10.1038/tp.2013.123.

[239] Li J, Liu J, Zhao L, Ma Y, Jia M, Lu T, Ruan Y, Li Q, Yue W, Zhang D, Wang L. Association study between genes in Reelin signaling pathway and autism identifies DAB1 as a susceptibility gene in a Chinese Han population. Prog Neuropsychopharmacol Biol Psychiatry 2013; 44:226–32. doi:10.1016/j.pnpbp.2013.01.004.

[240] Ozonoff S, Iosif A-M, Baguio F, Cook IC, Hill MM, Hutman T, Rogers SJ, Rozga A, Sangha S, Sigman M, Steinfeld MB, Young GS. A prospective study of the emergence of early behavioral signs of autism. J Am Acad Child Adolesc Psychiatry 2010; 49:256–66.e1–2.

[241] Sheinkopf SJ, Iverson JM, Rinaldi ML, Lester BM. Atypical cry acoustics in 6-month-old infants at risk for autism spectrum disorder. Autism Res 2012; 5:331–9. doi:10.1002/aur.1244.

[242] Paul R, Fuerst Y, Ramsay G, Chawarska K, Klin A. Out of the mouths of babes: vocal production in infant siblings of children with ASD. J Child Psychol Psychiatry 2011; 52:588–98. doi:10.1111/j.1469-7610.2010.02332.x.

[243] Ey E, Leblond CS, Bourgeron T. Behavioral profiles of mouse models for autism spectrum disorders. Autism Res 2011; 4:5–16. doi:10.1002/aur.175.

[244] Crawley JN. Designing mouse behavioral tasks relevant to autistic-like behaviors. Ment Retard Dev Disabil Res Rev 2004; 10:248–58. doi:10.1002/mrdd.20039.

[245] Noirot E. Ultra-sounds in young rodents. I. Changes with age in albino mice. Anim Behav 1966; 14:459–62.

[246] Bowers JM, Perez-Pouchoulen M, Edwards NS, McCarthy MM. Foxp2 mediates sex differences in ultrasonic vocalization by rat pups and directs order of maternal retrieval. J Neurosci 2013; 33:3276–83. doi:10.1523/jneurosci.0425-12.2013.

[247] Takahashi T, Okabe S, Broin PÓ, Nishi A, Ye K, Beckert MV, Izumi T, Machida A, Kang G, Abe S, Pena JL, Golden A, Kikusui T, Hiroi N. Structure and function of

neonatal social communication in a genetic mouse model of autism. Mol Psychiatry 2016; 21:1208–14. doi:10.1038/mp.2015.190.

[248] Chadman KK, Gong S, Scattoni ML, Boltuck SE, Gandhy SU, Heintz N, Crawley JN. Minimal aberrant behavioral phenotypes of neuroligin-3 R451C knockin mice. Autism Res 2008; 1:147–58. doi:10.1002/aur.22.

[249] Scattoni ML, McFarlane HG, Zhodzishsky V, Caldwell HK, Young WS, Ricceri L, Crawley JN. Reduced ultrasonic vocalizations in vasopressin 1b knockout mice. Behav Brain Res 2008; 187:371–8. doi:10.1016/j.bbr.2007.09.034.

[250] Fujita E, Tanabe Y, Shiota A, Ueda M, Suwa K, Momoi MY, Momoi T. Ultrasonic vocalization impairment of Foxp2 (R552H) knockin mice related to speech-language disorder and abnormality of Purkinje cells. Proc Natl Acad Sci USA 2008; 105:3117–22. doi:10.1073/pnas.0712298105.

[251] Shu W, Cho JY, Jiang Y, Zhang M, Weisz D, Elder GA, Schmeidler J, De Gasperi R, Sosa MAG, Rabidou D, Santucci AC, Perl D, Morrisey E, Buxbaum JD. Altered ultrasonic vocalization in mice with a disruption in the Foxp2 gene. Proc Natl Acad Sci USA 2005; 102:9643–8. doi:10.1073/pnas.0503739102.

[252] Young DM, Schenk AK, Yang S-B, Jan YN, Jan LY. Altered ultrasonic vocalizations in a tuberous sclerosis mouse model of autism. Proc Natl Acad Sci USA 2010; 107:11074–9. doi:10.1073/pnas.1005620107.

[253] Scearce-Levie K, Roberson ED, Gerstein H, Cholfin JA, Mandiyan VS, Shah NM, Rubenstein JLR, Mucke L. Abnormal social behaviors in mice lacking Fgf17. Genes Brain Behav 2008; 7:344–54. doi:10.1111/j.1601-183X.2007.00357.x.

[254] Winslow JT, Hearn EF, Ferguson J, Young LJ, Matzuk MM, Insel TR. Infant vocalization, adult aggression, and fear behavior of an oxytocin null mutant mouse. Horm Behav 2000; 37:145–55. doi:10.1006/hbeh.1999.1566.

[255] Folsom TD, Fatemi SH. The involvement of Reelin in neurodevelopmental disorders. Neuropharmacology 2013; 68:122–35. doi:10.1016/j.neuropharm.2012.08.015.

[256] Biamonte F, Assenza G, Marino R, D'Amelio M, Panteri R, Caruso D, Scurati S, Yague JG, Garcia-Segura LM, Cesa R, Strata P, Melcangi RC, Keller F. Interactions between neuroactive steroids and reelin haploinsufficiency in Purkinje cell survival. Neurobiol Dis 2009; 36:103–15. doi:10.1016/j.nbd.2009.07.001.

[257] Liu WS, Pesold C, Rodriguez MA, Carboni G, Auta J, Lacor P, Larson J, Condie BG, Guidotti A, Costa E. Down-regulation of dendritic spine and glutamic acid decarboxylase 67 expressions in the reelin haploinsufficient heterozygous reeler mouse. Proc Natl Acad Sci USA 2001; 98:3477–82. doi:10.1073/pnas.051614698.

[258] Marrone MC, Marinelli S, Biamonte F, Keller F, Sgobio CA, Ammassari-Teule M, Bernardi G, Mercuri NB. Altered cortico-striatal synaptic plasticity and related behavioural impairments in reeler mice. Eur J Neurosci 2006; 24:2061–70. doi:10.1111/j.1460-9568.2006.05083.x.

[259] Ammassari-Teule M, Sgobio C, Biamonte F, Marrone C, Mercuri NB, Keller F. Reelin haploinsufficiency reduces the density of PV+ neurons in circumscribed regions of the striatum and selectively alters striatal-based behaviors. Psychopharmacology (Berl) 2009; 204:511–21. doi:10.1007/s00213-009-1483-x.

[260] Akbarian S, Huang H-S. Molecular and cellular mechanisms of altered GAD1/GAD67 expression in schizophrenia and related disorders. Brain Res Rev 2006; 52:293–304. doi:10.1016/j.brainresrev.2006.04.001.

[261] Kemper TL, Bauman ML. The contribution of neuropathologic studies to the understanding of autism. Neurol Clin 1993; 11:175–87.

[262] Courchesne E, Yeung-Courchesne R, Press GA, Hesselink JR, Jernigan TL. Hypoplasia of cerebellar vermal lobules VI and VII in autism. N Engl J Med 1988; 318:1349–54. doi:10.1056/NEJM198805263182102.

[263] Di Martino A, Kelly C, Grzadzinski R, Zuo X-N, Mennes M, Mairena MA, Lord C, Castellanos FX, Milham MP. Aberrant striatal functional connectivity in children with autism. Biol Psychiatry 2011; 69:847–56. doi:10.1016/j.biopsych.2010.10.029.

[264] Sears LL, Vest C, Mohamed S, Bailey J, Ranson BJ, Piven J. An MRI study of the basal ganglia in autism. Prog Neuropsychopharmacol Biol Psychiatry 1999; 23:613–24.

[265] Ognibene E, Adriani W, Macrì S, Laviola G. Neurobehavioural disorders in the infant reeler mouse model: interaction of genetic vulnerability and consequences of maternal separation. Behav Brain Res 2007; 177:142–9. doi:10.1016/j.bbr.2006.10.027.

[266] Laviola G, Adriani W, Gaudino C, Marino R, Keller F. Paradoxical effects of prenatal acetylcholinesterase blockade on neuro-behavioral development and drug-induced stereotypies in reeler mutant mice. Psychopharmacology (Berl) 2006; 187:331–44. doi:10.1007/s00213-006-0426-z.

[267] Ognibene E, Adriani W, Granstrem O, Pieretti S, Laviola G. Impulsivity-anxiety-related behavior and profiles of morphine-induced analgesia in heterozygous reeler mice. Brain Res 2007; 1131:173–80. doi:10.1016/j.brainres.2006.11.007.

[268] Mullen BR, Khialeeva E, Hoffman DB, Ghiani CA, Carpenter EM. Decreased reelin expression and organophosphate pesticide exposure alters mouse behaviour and brain morphology. Asn Neuro 2012; 5:e00106. doi:10.1042/AN20120060.

[269] Biamonte F, Latini L, Giorgi FS, Zingariello M, Marino R, De Luca R, D'Ilio S, Majorani C, Petrucci F, Violante N, Senofonte O, Molinari M, Keller F. Associations among exposure to methylmercury, reduced Reelin expression, and gender in the cerebellum of developing mice. Neurotoxicology 2014; 45:67–80. doi:10.1016/j.neuro.2014.09.006.

[270] Romano E, Michetti C, Caruso A, Laviola G, Scattoni ML. Characterization of neonatal vocal and motor repertoire of reelin mutant mice. PLoS ONE 2013; 8:e64407. doi:10.1371/journal.pone.0064407.

[271] Vernes SC, Spiteri E, Nicod J, Groszer M, Taylor JM, Davies KE, Geschwind DH, Fisher SE. High-throughput analysis of promoter occupancy reveals direct neural targets of FOXP2, a gene mutated in speech and language disorders. Am J Hum Genet 2007; 81:1232–50. doi:10.1086/522238.

[272] Pramatarova A, Chen K, Howell BW. A genetic interaction between the APP and Dab1 genes influences brain development. Mol Cell Neurosci 2008; 37:178–86.

[273] Abadesco AD, Cilluffo M, Yvone GM, Carpenter EM, Howell BW, Phelps PE. Novel Disabled-1-expressing neurons identified in adult brain and spinal cord. Eur J Neurosci 2014; 39:579–92. doi:10.1111/ejn.12416.

[274] Frykman PK, Brown MS, Yamamoto T, Goldstein JL, Herz J. Normal plasma lipoproteins and fertility in gene-targeted mice homozygous for a disruption in the gene encoding very low density lipoprotein receptor. Proc Natl Acad Sci USA 1995; 92:8453–7.

[275] Trommsdorff M, Gotthardt M, Hiesberger T, Shelton J, Stockinger W, Nimpf J, Hammer RE, Richardson JA, Herz J. Reeler/Disabled-like disruption of neuronal migration in knockout mice lacking the VLDL Receptor and ApoE Receptor 2. Cell 1999; 97:689–701.

[276] Peñagarikano O, Abrahams BS, Herman EI, Winden KD, Gdalyahu A, Dong H, Sonnenblick LI, Gruver R, Almajano J, Bragin A, Golshani P, Trachtenberg JT, Peles

E, Geschwind DH. Absence of CNTNAP2 leads to epilepsy, neuronal migration abnormalities, and core autism-related deficits. Cell 2011; 147:235–46. doi:10.1016/j.cell.2011.08.040.

[277] Condro MC, White SA. Recent advances in the genetics of vocal learning. Comp Cogn Behav Rev 2014; 9:75–98. doi:10.3819/ccbr.2014.90003.

[278] Chabout J, Sarkar A, Dunson DB, Jarvis ED. Male mice song syntax depends on social contexts and influences female preferences. Front Behav Neurosci 2015; 9:96. doi:10.3389/fnbeh.2015.00076.

[279] Hahn ME, Lavooy MJ. A review of the methods of studies on infant ultrasound production and maternal retrieval in small rodents. Behav Genet 2005; 35:31–52. doi:10.1007/s10519-004-0854-7.

[280] Hammerschmidt K, Whelan G, Eichele G, Fischer J. Mice lacking the cerebral cortex develop normal song: Insights into the foundations of vocal learning. Nature 2015; 5:8808. doi:10.1038/srep08808.

[281] Bolhuis JJ, Okanoya K, Scharff C. Twitter evolution: converging mechanisms in birdsong and human speech. Nat Rev Neurosci 2010; 11:747–59. doi:10.1038/nrn2931.