# Molecular Specializations of a Songbird Vocal Motor Region

By

Alexander A. Nevue

A Dissertation

Presented to the Neuroscience Graduate Program

and the Oregon Health and Science University

School of Medicine

In partial fulfillment of

the requirements for the degree of

Doctor of Philosophy

September 15, 2022

# Certificate of Approval

This is to certify that the PhD dissertation of

Alexander A. Nevue

has been approved

_____
Claudio Mello, MD, PhD, Mentor

_____
Laurence Trussell, PhD, Chair

_____
Hiroyuki Nakai, MD, PhD, Member

_____
Stephen David, PhD, Member

_____
Kevin Wright, PhD, Member

_____
Arpiar Saunders, PhD, Member

# Acknowledgements

I would like to thank my advisor Dr. Claudio Mello for giving me the intellectual freedom to explore my interests and become an independent thinker during my PhD. Thank you to former Mello lab members Samantha Friedrich, Peter Lovell, Alexandra Wilmington, and Morgan Wirthlin for help with my projects along the way. Thank you to Dr. Henrique von Gersdorff and Benjamin Zemel who I closely collaborated with on multiple studies during my time in graduate school. Thank you to Dr. Hiroyuki Nakai for your generosity with both your time and resources. My time spent in your lab inspired my interest in virology and future research directions. Thank you to the entire Nakai lab especially Samuel Huang and Anusha Sairavi for your help and patience.

Thank you to my dissertation advisory committee: Dr. Laurence Trussell, Dr. Stephen David, Dr. Kevin Wright, and Dr. Hiroyuki Nakai. I appreciate your feedback and guidance. Thank you to Dr. Arpiar Saunders for being on my exam committee. Thank you to Dr. Henryk Urbanski whose Neuroscience of Aging T32 program supported my stipend for two years. Thank you to Dr. Stephen Moore and the members of the MMG journal club which was one of my favorite parts of grad school.

Thank you to my family: Mom, Dad, Andrew, Logan, Burton, and Stanley who provided an immense amount of support during grad school. And thank you to Caitlynn De Preter because our long walks solved every problem I faced.

**Table of Contents**


*Introduction* ..... *6*

**1.1 The song system** ..... **7**

**1.2 Defining molecular specializations in the song system** ..... **9**

**1.3 Outstanding questions in molecular architecture of RA** ..... **13**

**1.4 Goals and summary of dissertation** ..... **15**

*Molecular specializations of deep cortical layer analogs in songbirds* ..... *20*

**Abstract** ..... **20**

**Introduction** ..... **21**

**Methods** ..... **24**

*Animals and tissue preparation*: ..... 24

*In situ hybridization:* ..... *25*

*In situ mapping and image analysis:* ..... *28*

Bioinformatics ..... 30

**Results** ..... **31**

**Discussion** ..... **36**

*Molecular specializations in a vocal-motor brain region at the cell type level* ..... *54*

**Abstract** ..... **54**

**Introduction** ..... **55**

**Results** ..... **58**

*Cell type composition of the intermediate arcopallium (AI)* ..... 58

*Molecular architecture of excitatory neurons in AI* ..... 59

*Inhibitory neuron subtypes in the AI* ..... 64

*Non-neuronal cell specializations* ..... 67

*Insights into RA physiology and function* ..... 69

*Interactive apps for data exploration* ..... 71

**Discussion** ..... **72**

**Methods** ..... **79**

Animals ..... 79

Single nuclei isolation and sequencing ..... 79

Bulk RNA-sequencing ..... 81

*In situ* hybridization ..... 82

Analysis of in situ hybridization images ..... 82

Biophysical characterization of GABAergic neurons ..... 83

Interactive application development ..... 84

*Sodium and potassium ion channel specializations in the songbird intermediate arcopallium* ..... *104*

**Abstract:** ..... **104**

**Introduction: ..... 105**

**Results: ..... 107**

*Increase of Navβ4 in RA during vocal development* ..... 107

*Kv3.1 is the TEA-hypersensitive ion channel subunit preferentially expressed in RAPNs* ..... 110

**Discussion ..... 114**

**Methods ..... 116**

Animal subjects ..... 116

In situ hybridization ..... 116

Slice preparation for electrophysiology experiments ..... 118

***Naturally occurring gene knockout attenuates AAV transduction in songbirds ..... 127***

**Abstract ..... 127**

**Introduction ..... 128**

**Results ..... 128**

**Discussion ..... 131**

**Methods ..... 132**

*Genomics* ..... 132

*In situ hybridization* ..... 133

*Cell Culture* ..... 134

*Viral production* ..... 134

*Stereotactic injections and tissue processing* ..... 134

*AAV Barcode-Seq* ..... 135

***Summary and Discussion ..... 139***

**6.1 Defining molecular specializations of RA ..... 139**

**6.2 Impacts to AAV biology ..... 143**

**6.3 Future Directions ..... 146**

**6.4 Conclusion ..... 148**

# Chapter 1

# Introduction

The human brain executes complex motor behaviors such as speech or running or juggling, but the specific cell types, gene expression patterns, and physiological profiles that enable these behaviors are largely unknown. As children, humans learn how to speak, a highly complex motor behavior, through imitation – a process referred to as vocal learning. Acquisition of language can be disrupted in a variety of ways including hearing impairment (e.g., congenital deafness)[1] and neurodevelopmental disorders such as autism spectrum disorder[2]. The critical nature of spoken language in our culture and the prevalence of disorders that can affect speech makes it a critical behavior to study. However, studies in humans are limited due to the inaccessibility of the human brain and lack of experimental manipulations. Biomedical research therefore utilizes a variety of organisms, from cell lines to non-human primates, to model specific feature. Studying acquisition of language is complicated by the fact that vocal imitation is extremely rare among animals[3]. Commonly used laboratory animals such as mice do not learn vocalizations by

*Figure 1.1. Sexual dimorphism of zebra finches. A female (right) with two males (center, left). The adult male zebra finch can be differentiated by feather patterns including an orange cheek patch and black and white "zebra" stripes on their neck.*

imitation[4, 5]. A deaf mouse will squeak just the same as a normal hearing mouse. Neuroethology is a subfield of neuroscience that exploits lineage-specific behaviors of animals in an attempt to gain insights into how the brain functions. During my PhD, I utilized the zebra finch, which produces a learned song, as a model organism to study the complex motor behavior vocal imitation.

# 1.1 The song system

The zebra finch is one of approximately 4,000 songbird species which learn a song via imitation. Zebra finches are desirable in a laboratory environment because they are small and easy to breed and maintain[6]. Male zebra finches, who exhibit a sexual dimorphism in their physical appearance (Fig. 1.1) and in their behavior[7, 8], sing a highly stereotyped song during courtship[9]. This courtship behavior is supported by dedicated brain circuitry[10, 11] (Fig. 1.2). The gross brain anatomy of the avian brain differs from the mammalian architecture[10]. While subcortical circuits and structures are similar, the two lineages diverge heavily with their cortical architecture. Instead of having a layered cortex like mammals, birds instead have a pallium comprised of nuclei[12]. Despite the different architecture, these nuclei are connected in similar ways as the canonical cortical microcircuit[13, 14] where individual spatially segregated nuclei function as cortical layers. Songbirds have two main brain circuits for the learning and production of song, together forming what is called the song system. First, a cortico-basal ganglia-thalamocortical loop called the anterior forebrain pathway, which is critical for learning a song[15, 16]. Brain regions in this loop inject variability into the developing song and help crystallize the stereotyped adult song[17]. The other pathway, which has been the focus of my PhD, is called the vocal motor pathway. This pathway is directly involved in the

production of song and when disrupted, causes the bird to be unable to produce song[7]. It starts with a premotor input from a nucleus called HVC which projects into the robust nucleus of the arcopallium (RA). HVC neurons fire sparsely to provide precise timing input to RA[18]. The role of HVC being critical for timing was confirmed with a brain cooling experiment where decreasing the temperature of HVC while the bird was singing elongated the duration of the song[19]. This is in stark contrast to RA neurons which fire bursts of action potential throughout the song[18]. RA neurons make a direct projection to the brainstem motor neurons that control the vocal organ in birds [20] (avian syrinx is analogous to mammalian larynx). This direct projection from RA to the brainstem is unique to animals that produce learned vocalizations and allow for more control than if the projection went through the midbrain. Therefore, in function, RA is analogous to deep layer pyramidal neurons in the mammalian motor cortex.

RA was first described in 1976 in a canary [7]. The seminal but somewhat crude experiment lesioned RA which prevented the bird from singing, showing the necessity of RA in the song system. In the decades since, songbird research has focused on describing how RA and the other song nuclei behave at the molecular, cellular, and systems level. Notable findings include the drastic sexual dimorphism in RA size between male and female zebra finches[7], and in vivo recordings showing that RA fires highly stereotyped bursts of action potentials throughout a song suggesting RA controls

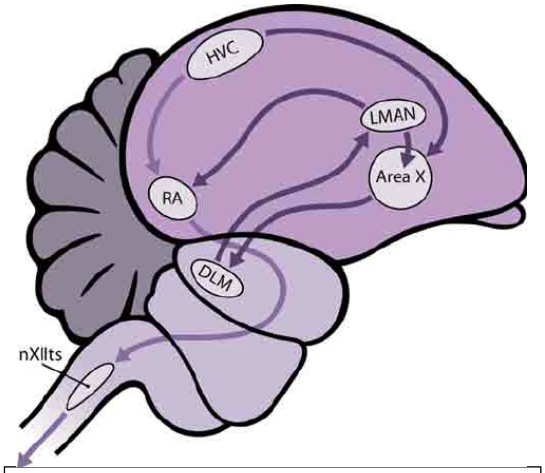



*Figure 1.2. The song system of a zebra finch. The song system has two main components: 1) a cortico-striatial-thalamocortical loop which projects into 2) the motor output pathway which consists of a premotor HVC input into the extratelencephalic RA to brainstem.*

frequency modulation of song [18]. What is less known however, is the molecular architecture of RA that enables its unique abilities.

## 1.2 Defining molecular specializations in the song system

A long-term goal of the Mello lab is to characterize the gene expression specializations within the song system. A large majority of this effort has been through spatially defining gene expression specializations via in situ hybridization, where a nucleic acid probe binds to (or hybridizes) a specific mRNA within a tissue [21]. Using a variety of detection protocols then allows for the visualization of which cells in which brain region express a gene of interest. By comparing the expression of a gene within a region of interest, for example RA, to the brain regions that surround RA, one can make a more definitive case for a "specialization." We can then use these specializations to generate hypotheses about how songbirds learn and produce song [22]. These efforts have yielded two main categories of gene expression specializations (also referred to as gene markers). First, positive markers where the region of interest expresses more of the gene compared to the surrounding region. Second, negative markers, where the region of interest does not express the gene and the surrounding region does. While simplistic, the importance of using context to define gene expression specializations cannot be understated. Molecular context, the term I use for a spatially aware molecular comparison, became a theme throughout my PhD and changed how we thought about the organization and evolution of the song system.

To illustrate the importance of molecular context, imagine a scenario where you only have access to a small 0.5x0.5mm square piece of RA tissue with in situ hybridizations for two

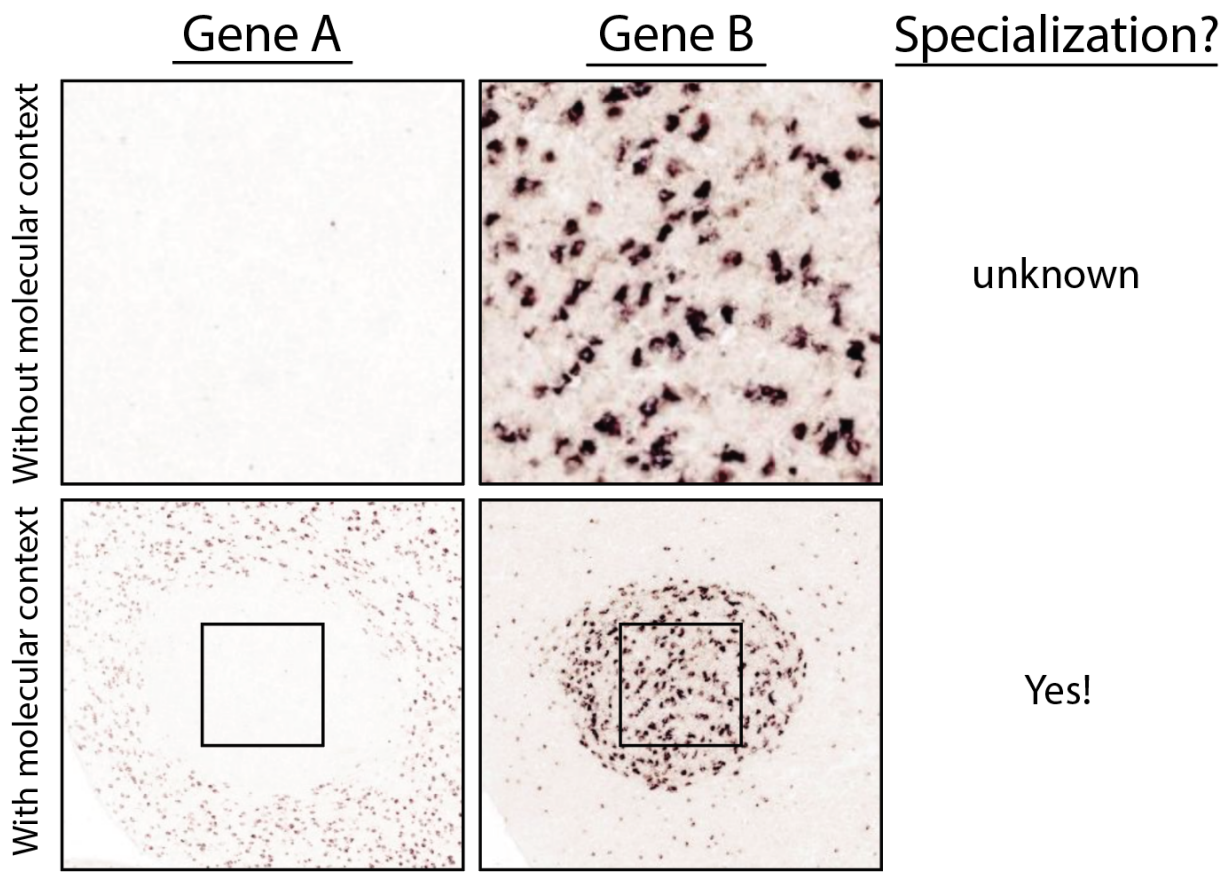



*Figure 1.3. Molecular context. When analyzing only the region of interest, molecular context (i.e. the gene expression profile of the surrounding region) is missing (top). Information such as present or absence of expression can be ascertained but no information on relative expression is known. With molecular context (bottom) assessing if the gene expression of the region of interest is unique is now possible.*

different genes. Your goal is to determine what makes RA unique. Gene A is completely blank, no cells are labeled. Gene B has many cells labeled. What can you conclude about RA? You only know Gene A is absent and Gene B is present. Is this enough to determine a specialization? For all we know Gene A could be a gene not expressed in the brain and Gene B expressed in every cell in the brain. This analysis is not very helpful towards your goal of defining what is unique to RA. Now instead, you have in situ hybridizations on the entire brain section that contains RA. With molecular context you now learned that Gene A is absent in RA but expressed everywhere

else in the brain and Gene B is highly expressed in RA but low in the surrounding region. You can now generate hypotheses about how these genes impact the physiology of RA (Fig. 1.3).

The idea of molecular context can be applied to performing RNA sequencing (RNAseq), where the sequence of individual mRNA molecules is read and mapped onto a reference genome to identify the gene, in a given brain region [23]. You will find out the genes expressed in your region of interest but without a comparison region you are unable to conclude anything regarding the specificity of gene expression. A gene expressed at the same levels in the same types of cells throughout the brain likely does not have an impact on the specific role of your brain region of interest. Utilizing a comparison region in an RNAseq experiment allows for a differential screen, which can generate new markers. These differential screens have been utilized in the song system and have produced numerous examples of putative specializations [24, 25]. More recently, cell type information can also be gathered from RNAseq experiments [26-28]. The most common high throughput single cell RNAseq (scRNAseq) method uses oil droplets and DNA barcodes to sequester and label mRNA from individual cells with unique barcodes[26, 27]. Now, not only are differential screens possible to identify unique gene expression but it is also possible to determine how specific cell classes differ between brain regions.

Prior to the invention of genome-wide (or near genome-wide) transcriptomics assays, markers of song nuclei were determined by cloning and probing a gene in tissue using a technique called in situ hybridization (Fig. 1.4). This is a powerful technique to analyze the spatial distribution of a gene, however, this strategy of cloning and examining the expression of individual genes can be low yield and time consuming for screening. High-throughput methods

to profile the transcriptome of a brain region is species agnostic but importantly requires a quality reference genome. The zebra finch genome was first sequenced in 2010 [29] and allowed for studies into the link between genetics and behavior [30]. More recently, updated genomes using long read PacBio technology have been generated and released which provide better coverage for genomic analysis [31]. In fact, some avian genomes are among the most complete out of any reference genome which enhances their contributions as model organisms [32].

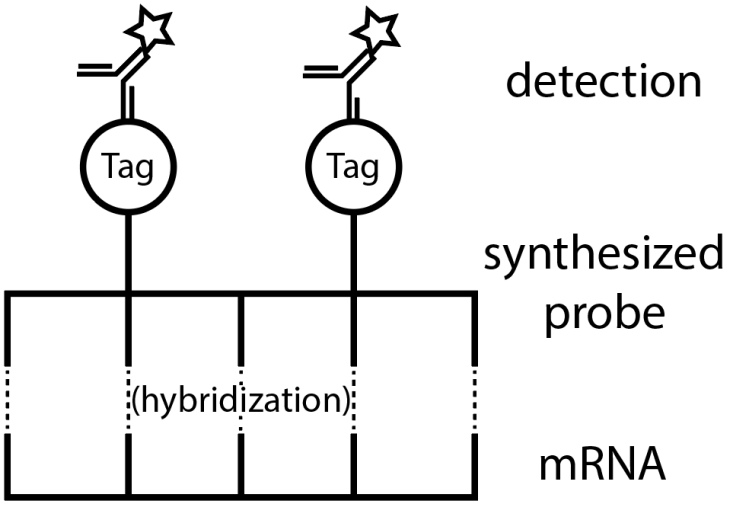



*Figure 1.4. In situ hybridization. A probe made from DNA or RNA is synthesized to bind to a specific mRNA species. The synthesized probe is tagged with a molecule that can then be detected and visualized, often with an antibody.*

One landmark study in the vocal learning field performed a differential microarray expression analysis to compare vocal regions in songbirds and humans and found similar gene expression profiles in the vocal regions of the two species [24]. Despite both humans and songbirds exhibiting imitative vocal behavior, birds and mammals share a common ancestor 300 million years ago. This behavior is an example of convergent evolution where separate groups evolve similar features that the common ancestor does not have. A common example of convergent evolution is the evolution of flight in bats, birds, and insects – none of whose ancestor had wings. From the differential microarray, it was found that RA and the analogous region in human (laryngeal motor cortex – LMC [33]) exhibited similar gene expression for a set of approximately 50 genes. This finding suggests that not only is the behavior of vocal learning

an example of convergent evolution but that there may be a convergence in the gene expression of these two areas. The vocal production regions may require a specific molecular architecture to make direct connections to the brainstem or produce fast, precise bursts of action potentials. This study was one of the first to truly make a neural link (as opposed to behavioral) between birds and mammals. However, multiple issues presented itself related to the full conclusions of this study in terms of pure RA specializations as well as comparisons to mammals.

# 1.3 Outstanding questions in molecular architecture of RA

The Pfenning et al. [24] study was seminal in the field as it was the first to compare avian and mammalian brains beyond a select few genes. Shortly after I joined the Mello lab, the lab published an atlas of the zebra finch arcopallium in which RA resides [34]. They found some genes in a frontal section appeared to have continuous gene expression with RA and a neighboring brain region (Fig. 1.5). This neighboring brain region is called the dorsal intermediate arcopallium (AId) and while far less studied than RA is thought to be a motor region responsible for non-vocal movements such as wing flapping or hopping [35]. AId is thought to be analogous to deep cortical layer neurons of the mammalian cortex based

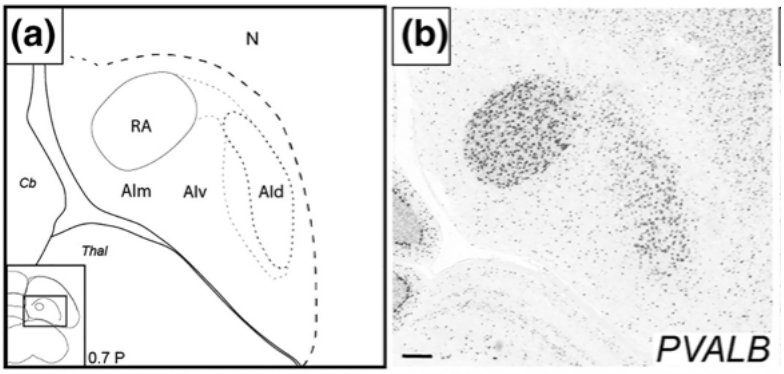



*Figure 1.5. The intermediate arcopallium of the zebra finch. A) a schematic of a frontal section outlining RA and AId. B) an in situ hybridization for PVALB showing similar and near-continuous expression between RA and AId. Adapted from Mello et al., 2019 JCN.*

on connectivity [36] and transcription factor expression [37]. My first project in lab, which ended up turning into Aim 1 of my dissertation, was to further compare the gene expression relationship between RA and AId and I started with the set of genes identified in Pfenning et al. [24] as they represented some of the most critical genes to songbird vocal biology. I found that many of these genes were not specific to RA and also were molecular markers of AId [38]. This was a very surprising finding because not only were these genes supposed to be unique to RA but they were supposed to be the link between bird and human vocal behavior. I found an explanation for this discrepancy in the methodology of the differential microarray screen. The comparison region for RA that was used was not AId but instead an adjacent auditory region called the ventral intermediate arcopallium [39]. Instead of identifying vocal motor specializations by comparing RA to a motor region (like AId) this screen compared a vocal to auditory region and therefore resulted in identifying mostly general motor specializations. This illustrates the importance of molecular context and choosing the best comparison region for differential screens.

The true molecular specializations of RA remained unknown. Despite nearly five decades of RA research some of the most basic descriptions of RA have remained elusive because of the complex architecture of the songbird brain. Determining the molecular specializations of RA is critical in learning how RA is contributes to the production of learned vocalizations.

## 1.4 Goals and summary of dissertation

Prior to defining molecular specializations, I realized it was important to identify the proper comparison region for RA. I built off of previous findings and showed that RA and AId exhibit striking similarities in gene expression [34]. The proximity and similar gene expression of RA and AId points to these two brain regions sharing similar functions [35]. However, this similarity alone is not sufficient to conclude that AId is the proper comparison region for molecular context. In **Aim 1** of my dissertation, I analyzed the expression of 162 genes, in a dozen subdivisions of the arcopallium, to determine molecular relationships among the various arcopallial subdivisions. I found that overwhelmingly RA and AId had the highest percentage of shared gene expression. Not only was this an important finding that influenced future experiments in the lab but it gave insights into the function and evolution of RA. My spatial analysis defined, for the first time, a subset of genes with truly unique expression in RA. Furthermore, these results provided support for a hypothesis for how the song system evolved. While only vocal learning birds (songbirds, parrots, hummingbirds) possess an RA, other birds (e.g. chicken, ducks) lack song nuclei while maintaining a similar arcopallial structure. The motor theory for vocal learning origin states that the song system evolved from nearby pre-existing motor regions [35, 40]. With respect to RA, that would be the adjacent motor region AId. A high similarity in gene expression between the two regions supports that RA evolved as a specialization of the pre-existing AId. I further showed support for this hypothesis by using zebra finch RA/AId markers in suboscines – a group of birds closely related to songbirds but that are not vocal learners [41]. In these brains I found that the suboscine AId was labeled with

these markers supporting that AId is an evolutionary conserved region with conserved gene expression. I used these findings to influence my second aim.

Now that I had identified that RA and AId share this close molecular relationship, I wanted to design an experiment to determine true molecular specializations of RA. One option would be a differential RNAseq screen where I would individually dissect RA and AId and determine how they differ in gene expression. This is a perfectly valid experiment but lacks cell type resolution. In bulk RNAseq experiments the tissue is homogenized, cells are lysed, and then RNA is extracted. After analysis, when an enrichment of Gene A is observed, it is not possible to identify the identity of the cell (neuronal vs non-neuronal, excitatory vs inhibitory neuron etc) from which that RNA originated. Using single cell or single nucleus RNA sequencing (sc/snRNAseq) all sequenced RNA molecules from a given cell are labeled with the same molecular barcode allowing for molecular specialization conclusions with cell type specificity. Cell type information is highly desirable in RA because excitatory neurons represent the projection neurons that are directly involved with producing song. Despite their critical nature in song, the molecular specializations of RA projection neurons are unknown. Concluding that given molecular specializations are restricted to RA projection neurons would be impactful for interpreting a potential role in vocal production. Beyond excitatory neurons, little is known about

Using known gene expression patterns from our spatial in situ atlas ZEBrA (Zebra Finch Brain Atlas) [22] and from Aim 1 [38], I was confident that RA projection neurons exhibited multiple markers (e.g. GABRE, KCNC1) and therefore that I would be able to differentiate them molecularly from other cell types. Importantly, I would be able to differentiate RA projection

neurons from other excitatory neurons in the intermediate arcopallium including AId. In **Aim 2** I used snRNAseq to identify molecular specializations in RA with cell type specificity. I explored multiple vignettes in this aim including describing non-neuronal specializations, GABAergic neuron distribution, and critical insights into RA projection neuron specializations. My strategy of using cell type resolution with proper molecular context allowed for the identification of RA gene expression specializations that previous studies have missed. RA projection neuron specializations included growth factors and testosterone synthesis genes that explain previous work in the field [7, 8, 42, 43]. I integrated the findings in this study with a bulk RNAseq dataset from male and female RA during vocal development [44]. I found that markers associated with inhibitory neurons decrease in RA during vocal development and non-neuronal markers increase in expression. Finally, I designed an interactive application and integrated it with the ZEBrA atlas so that others have access to the data from this study.

In **Aim 3** I collaborated with a lab specializing in slice electrophysiology to associate RA markers with cellular properties of RA projection neurons. RA projection neurons are known to fire extremely fast to enable vocalizations [18]. While many molecular specializations of the song system have been identified, few have confirmed roles in the zebra finch brain. First, I described a molecular switch of sodium ion channel subunits in RA [45]. Prior to song learning, RA has high expression of the sodium channel subunit SCN3B and low expression of SCN4B. During song learning I found that the expression of SCN3B decreased to being completely absent in the adult, yet the expression of SCN4B increased and turned into a positive marker of adult RA. SCN4B, which encodes for the NavB4 protein, is associated with a resurgent sodium current that assists neurons in firing another action potential. The developmental expression of

SCN4B was highly correlated with the developmental emergence of a resurgent sodium current. Second, I described the differential expression of potassium ion channel subunits in RA compared to AId. Patch clamp recordings showed that the action potential waveforms in RA are faster than in AId, suggesting voltage gated potassium channels may differ in expression between the two regions. My molecular analysis pointed towards KCNC1, the gene that encodes the protein Kv3.1, as the likely contributor which was corroborated using pharmacology. These two studies offer insights into the function of RA projection neurons and how they are able to fire rapid action potentials.

The goal of my first two aims was to identify true molecular markers of RA with the future goal to be characterizing how each of those molecular specializations contribute to song production. An example of how those specializations could be characterized was illustrated in Aim 3. Another way to gain insights into the role of a gene is to manipulate its expression. A common way systems neuroscientists manipulate the brain is with viruses, such as adeno-associated virus (AAV) [46]. AAVs enter the cell by hijacking cellular proteins which traffic the capsid to the nucleus where it is uncoated and the encoded genes are transcribed [47]. Using AAV, I could express negative markers in RA excitatory neurons to determine why RA has downregulated them or conversely, knockdown the expression of positive markers. In both cases measuring the effect on song could give insights into the function of the gene in RA. Unfortunately for songbird researchers, there has been very limited successes using viruses including AAVs in the songbird brain which has stunted discovery. Across multiple AAV serotypes and even other viruses, the transduction efficiency appears to be less than 10% of cells [48-50]. The goal of **Aim 4** was to determine which of the commonly used AAVs works best

in the songbird brain and to determine why AAVs are less efficient in the songbird brain compared to mammals. Using a DNA barcoded virus strategy [51], I found that AAV9 exhibited the highest transduction efficiency in RA as well as the zebra finch striatum. To investigate possible reasons for poor transduction efficiency I looked to a recent study describing putative AAV entry factors [52]. Analyzing the zebra finch genome, I found that the gene GPR108, one of the most critical genes in AAV transduction [53], is absent in the zebra finch genome suggesting a mechanism for lower AAV transduction compared to mammals. The results of this aim point to ways to utilize viruses for gene manipulation in the songbird and also the complex nature of virus-host evolution.

# Chapter 2

# Molecular specializations of deep cortical layer analogs in songbirds

This chapter was published as *Nevue AA, Lovell PV, Wirthlin M, Mello CV. Molecular specializations of deep cortical layer analogs in songbirds. Scientific Reports. 2020. 10(1),1-14.* A.A.N., P.V.L., M.W., C.V.M. conceived the study. A.A.N. performed the experiments and data analysis. A.A.N. and C.V.M. wrote the paper.

**Abstract**

How the evolution of complex behavioral traits is associated with the emergence of novel brain pathways is largely unknown. Songbirds, like humans, learn vocalizations via tutor imitation and possess a specialized brain circuitry to support this behavior. In a comprehensive *in situ* hybridization effort, we show that the zebra finch vocal robust nucleus of the arcopallium (RA) shares numerous markers (e.g. *SNCA*, *PVALB*) with the adjacent dorsal intermediate arcopallium (AId), an avian analog of mammalian deep cortical layers with involvement in motor function. We also identify markers truly unique to RA and thus likely linked to modulation of vocal motor function (e.g. *KCNC1, GABRE*), including a subset of the known shared markers between RA and human laryngeal motor cortex (e.g. *SLIT1*, *RTN4R*, *LINGO1*, *PLXNC1*). The data provide novel insights into molecular features unique to vocal learning circuits, and lend support for the motor theory for vocal learning origin.

## Introduction

An in-depth understanding of how the brain controls learned behaviors and how these behaviors arise in specific animal lineages requires detailed knowledge of the molecular organization of the underlying circuits. Songbirds offer an excellent model for investigating these questions. Their vocal circuitry has been extensively studied, and consists of interconnected pallial, basal ganglia, and thalamic components that control the production and acquisition of learned vocalizations. As is typical of birds, the pallial (cortical-like) areas consist of discrete nuclei, in contrast to the layered cortex of mammals [10, 22, 54]. The songbird vocal circuitry can be subdivided into a direct vocal-motor pathway, necessary for song production, and an anterior pathway, involved in vocal learning and adult vocal plasticity [6, 11, 55, 56]. Discrete nuclei of both pathways have also been identified in the other vocal learning avian groups (i.e. parrots [57-59] and hummingbirds [41, 60]) but are absent or rudimentary in vocal non-learning birds [41]. Notably, vocal learning behavior provides an important basis for spoken language in humans [61, 62] and the related circuits in vocal learning birds and humans share remarkable convergent molecular specializations [24]. In contrast, with the possible exception of a few other mammalian groups such as bats and cetaceans [63, 64], the occurrence of vocal learning and related circuitry is quite rare among vertebrates, and seems absent or only rudimentary in rodents and non-human primates [4, 65-67].

Despite considerable knowledge on the anatomical, physiological, and molecular properties of the songbird vocal circuitry, especially in zebra finches, our understanding of the evolution of

these anatomically and functionally distinct vocal nuclei is limited. The close proximity of the vocal nuclei to auditory areas has led to the hypothesis that these circuits may have evolved from circuits involved in song perceptual processing [68-70]. Alternatively, based on their close proximity to areas thought to be involved in motor control, it has been proposed that vocal nuclei evolved as specialized expansions of preexisting motor regions [35, 40, 71]. Progress towards testing these hypotheses has remained limited, however, possibly because most effort has focused on characterizing the unique properties of the vocal circuitry rather than on how they relate to other brain areas.

The robust nucleus of the arcopallium (RA) is a particularly prominent and extensively studied vocal nucleus in zebra finches. It is the major forebrain vocal output nucleus and is thought to encode important acoustic features of finch song motifs [18, 72, 73]. The RA is considered part of the intermediate arcopallium, which is the major source of descending output from the avian telencephalon [36]. The arcopallium, more broadly, is thought to contain the avian analogue of the deep layers of the mammalian sensory and motor cortices based on similarities in their projection patterns [70, 74], neuronal activation[35, 75], and transcriptional profiles [34, 37, 76, 77], but may also contain the avian equivalent of pallial parts of the mammalian amygdala [34, 78, 79]. Avian analogs of RA are found in other birds that evolved vocal learning [57, 59, 60], but are thought to be absent in vocal non-learning birds based on cytoarchitectonics and molecular criteria[80]. Transcriptomics studies [24, 25, 81] have identified several hundred differentially expressed genes in RA compared to the adjacent ventral intermediate arcopallium. Of these, a subset were described as molecular specializations shared with the

analogous nuclei in other vocal learner birds and the laryngeal representation of the primary motor cortex (LMC) in humans [24], suggesting that a shared gene network may have convergently evolved across different vocal learning systems. Notably, however, an extensive examination of the brain expression patterns of shared RA and LMC markers has not yet been performed.

We have previously described the differential expression of a small set of genes in both RA and the adjacent dorsal intermediate arcopallium (AId; referred to as LAI, Ad, and AI in previous studies), as well as the sharp borders of *SCN3B* expression for both RA and AId [34]. These observations are consistent with previous indications of similar connectivity between RA and AId. For example, RA and AId receive parallel input from the nidopallium [82, 83] and also send distinct but parallel projections to the brainstem that can be considered analogous to those in the cortico-bulbar tract in mammals. Other studies suggest common motor control functions of RA and AId. For example, AId shows immediate early gene expression after movements such as wing flapping or hopping [35, 75], analogous to RA being active during song production [18, 72, 84]. Furthermore, while lesions to RA result in severe song deficits, birds show marked motor deficits including akinesia and immobility in large lesions that primarily include AId [39]. These observations are consistent with AId being involved in somatic motor control [35], though other studies suggest a role in vocal learning [85]. It has been previously suggested that AId might be broadly present in birds, regardless of vocal learning, and that RA may have originated as an expansion and specialization of AId. Nonetheless, AId remains poorly defined, and our knowledge of its molecular organization is limited. A closer comparison of gene expression

patters in RA and AId is also needed to more clearly identify features unique to the vocal circuitry.

Our main goals were to improve our understanding of the molecular organization of AId in comparison with RA, and to better define molecular properties unique to RA. Using *in situ* hybridization for markers with sharp expression boundaries, we first generated a more precise definition of AId in adult male zebra finches. We then conducted extensive analysis to distinguish molecular features common to RA and AId from those unique to RA. We also were able to identify AId in pre-song juvenile males and non-singing females, as well as in two suboscine species, a sister taxa to songbirds generally thought to lack vocal learning and/or related forebrain vocal nuclei. Our data provide substantial further support for a close molecular similarity between RA and AId, as well as a more in-depth definition of features unique to RA and the vocal control circuitry. They suggest that AId may represent a subdivision of the arcopallium that existed prior to the emergence of vocal learning circuits in birds, and are consistent with the hypothesis that RA may have evolved as a specialization of AId.

## Methods

*Animals and tissue preparation*:

All procedures involving live animals were approved by the OHSU Institutional Animal Care and Use Committee and are in accordance with NIH guidelines. Adult (n=11) and fully fledged 20-day post-hatch (dph) (n=2) male, and adult female (n=2) zebra finches (*Taeniopygia guttata*) were obtained from our colony or purchased from a local breeder. Adult finches were isolated

in sound dampening chambers overnight and sacrificed by decapitation the next morning prior to lights on to minimize the potential confounds of singing and auditory stimulation on activity dependent changes in gene expression. To minimize possible adverse effects of stress, juvenile males were not sound isolated, and instead were removed directly from the aviary and sacrificed by decapitation a few minutes after lights on. Juvenile males could usually be identified by plumage, however we also confirmed sex by gonadal inspection. For all birds, immediately after sacrifice, their brains were dissected and blocked in either the sagittal plane (n= 3 adult male brains), or in the frontal plane (all other brains) at a level just rostral to the tectum, and frozen in Tissue-tek (Sakura-Finetek) in a dry ice/isopropyl alcohol slurry. Brains were sectioned on a Leica CM1850 cryostat at 10 µm thickness and mounted on charged microscope slides (Superfrost plus; Fisher Scientific). Sections were post fixed for 5 min at room temperature in a solution containing 3% paraformaldehyde in phosphate buffered saline (PBS), washed twice in PBS, dehydrated in an ethanol series, and stored at -80°C until use.

We also processed brains from a Willis's antbird (n=1 male; *Cercomacroides laeta*) and a Straight-billed woodcreeper (n=1 male; *Dendroplex picus*). These birds had been captured at field sites in the suburbs of Belém (Pará, Brazil) and deposited at the Emilio Goeldi Museum (Belem, PA, Brazil). Shortly after euthanasia, the brains were frozen and stored at -80$^{o}$C. These cryopreserved Museum samples were subsequently processed in the frontal plane as described above for zebra finches.

<u>In situ hybridization:</u>

We generated *in situ* hybridization data for 61 genes, including 46 genes previously identified as shared markers between avian RA analogs and human LMC [24] and/or highly differential markers of RA in zebra finches [25]. For each gene examined, we initially confirmed orthology between zebra finch and other clades by a combination of cross-species alignments and synteny verification with other birds (e.g. chicken) and bird outgroups (e.g. mouse, human), and non-avian sauropsids (e.g. anole), using UCSC's genome browser and the BLAT toolkit as previously detailed [22]. We then identified appropriate clones from the ESTIMA brain EST/cDNA library [86] for riboprobe synthesis. To maximize specificity, whenever possible we avoided clones containing protein-coding regions and selected those containing only or primarily the 3'-untranslated sequence. The clones selected for each gene were confirmed to align significantly to a single locus in the zebra finch genome. The clones for 36 of the genes included in the present study are listed in Supplemental Table 1; clone details for the remaining genes examined in the present study can be found in the Zebra finch Expression Brain Atlas (ZEBrA) website (www.zebrafinchatlas.org). Further details on our criteria and pipeline for clone selection for *in situs* has been previously described [22].

We followed protocols for riboprobe synthesis and purification, and *in* situ hybridization as described previously [21]. Briefly, selected cDNA clones were grown overnight, and plasmids were isolated (QIAprep spin Miniprep Kit), digested with BssHII, and purified (QIAquick PCR purification kit). Digoxygenin(DIG)-labeled antisense riboprobes were then synthesized using T3 RNA polymerase (Promega) and a DIG RNA labeling mix (Roche) for 2 hr at 37°C. Riboprobes were purified using Sephadex G-50 columns and stored at 20°C until use. Incubations with a no

probe negative control, or positive control probe for a gene with a known expression pattern (e.g. GAD2) were routinely included in hybridizations.

Prior to hybridization, slides were acetylated for 10 min in a solution containing 1.35% triethanolamine and 0.25% acetic anhydride. Slides were briefly washed in 2X SSPE (300 mM NaCl, 20 mM $NaH_2PO_4$-$H_2O$) and dehydrated in an ethanol series. A hybridization solution consisting of 50% formamide, 2X SSPE, 2 µg/µL tRNA, 1 µg/µL BSA, 1 µg/µL Poly A, and 4 µL DIG-labeled riboprobe in DEPC-treated $H_2O$ was prepared. Slides were coverslipped and hybridized in a mineral oil bath overnight at 65°C. The next day, the slides were washed in chloroform to remove the mineral oil and washed in SSPE to remove the coverslips. Slides were then washed in 50% formamide 2X SSPE solution followed by two 30 min washes in 0.1X SSPE at 65°C, agitated every 10 min. Following the high stringency washes, the sections were briefly permeabilized in TNT (100 mM Tris-HCl pH 7.4, 150 mM NaCl, 0.3% Triton X-100). Slides were then blocked in TNB (100 mM Tris-HCl pH 7.4, 150 mM NaCl, 0.36% w/v BSA, 1% skim milk) for 30 min in a humidified chamber at room temperature. Slides were then incubated in an alkaline phosphatase conjugated anti-DIG antibody (Roche, 1:600) in TNB for 2 hr in a humidified chamber at room temperature. Slides were then washed twice for 15 min in TMN (100 mM Tris-HCl, 150 mM NaCl, 5 mM $MgCl_2$) and incubated for 1-3 days in filtered BCIP/NBT Substrate Solution (PerkinElmer) at room temperature. After incubation, slides were rinsed in DI water, fixed in 3% paraformaldehyde, and washed again in DI water. Slides were then coverslipped with VectaMount permanent mounting medium (Vector).

All genes for which we generated *in situ* hybridization data were assessed in at least two brains of adult male zebra finches. Our *in situ* pipeline, consistent with that described for the ZEBrA database, consisted of an initial assessment of hybridization conditions and general expression pattern in one brain cut in the sagittal plane, and a final hybridization with sections containing RA and AId from another brain. *SCN3B* was run on all brains that were part of the study. For the AId reconstruction, the final hybridizations for *SCN3B* were run in frontal male and female brain series (2 brains each); specifically, every 10$^{th}$ slide (200 μm intervals) in the range that spans the arcopallium was stained for Nissl (cresyl violet), and adjacent slides were processed for *SCN3B in situ*. For the other 60 genes that were examined in frontal sections, effort was made to run the final hybridization at a level around the core region of RA and AId in at least one brain, including both the right and left hemispheres. The remaining 101 genes in this study were assessed in the sagittal plane only, most of them consisting of data that were already available on the ZEBrA website. A complete list of genes, and whether they were assessed in sagittal only or in both sagittal and frontal planes, can be found in Supplemental Table 2. Images from sagittal sections for most of the data generated here are being prepared for uploading to ZEBrA. Lastly, we note that the data in ZEBrA were generated using both left and right hemispheres, thus that database in its current form is not appropriate for evaluating possible hemispheric differences.

<u>In situ mapping and image analysis:</u>

For the AId reconstruction in the male and female frontal brains, the boundaries of major features such as section borders and laminae as seen in the Nissl-stained sections spanning the

arcopallium (every 10th slide, 200 µm intervals) were drawn using Neurolucida. In the adjacent hybridized sections, we then drew major section borders and the internal arcopallial boundaries that were defined by the differential expression of *SCN3B*. The resulting drawings were aligned to transverse sections[44].

For a qualitative analysis of markers of RA and/or AId, we visually examined the *in situ* patterns of 162 genes, including the patterns in frontal sections generated in this study as well as the available patterns in sagittal sections for another 116 genes classified as RA markers on the ZEBrA website. Our present analysis consisted of comparing gene expression in major arcopallial domains representing subdivisions of the anterior, medial, dorsal, posterior, intermediate, and ventral arcopallium, noting that to simplify the analysis we collapsed the previously defined 19 arcopallial subdivisions [34] into 12 major domains/subdomains for which we have numerous markers in ZEBrA. Genes that showed expression in a given arcopallial domain similar to the differential expression in RA were considered markers of both RA and that arcopallial domain. The comparison was not exclusive, so RA could share expression of a given marker with multiple arcopallial domains. The same approach was taken for the analysis of AId compared to other arcopallial subdomains. In this case, we started with the list of genes we identified as RA and AId markers (n=94 genes; see Table 1 in Results), noting that all known markers of AId are also markers of RA, as observed in the current study and in previous efforts [34]. For this reason, in the AId comparative gene expression analysis, RA was not included as a subdomain.

For a subset of genes examined in frontal sections (n=30), we measured relative expression levels within subdivisions of the intermediate arcopallium. This gene subset consisted of genes that were hybridized in closely adjacent sections, thus allowing a consistent evaluation across genes. Included were 7 genes qualitatively classified by visual inspection as markers of RA and AId, as well as 23 genes classified as markers of RA only. Using the FIJI distribution of NIH ImageJ, we performed average optical density measurements in 200 µm x 200 µm windows placed over RA and AId. We then calculated a relative expression ratio using the following ratio: $RA_{OD}/AId_{OD}$, where $RA_{OD}$ is the mean 8-bit grayscale value in RA for a given gene, and $AId_{OD}$ is the mean grayscale value in AId for that same gene. Thus, ratio values close to 1 correspond to genes with similar levels of expression in RA and AId, and values deviating from 1 correspond to genes that are differentially expressed in RA, either positive (ratio >1), or negative (ratio <1) compared to AId.

Bioinformatics

We used ConsensusPathDB to conduct a gene set over-representation analysis and identify individual genes associated with specific biological pathways. Our analysis was conducted using all available pathway databases, a minimum overlap with our input list of two genes and a p-value cutoff of 0.015. We analyzed separately two non-overlapping gene input lists (see Table 1 in Results), consisting of genes identified by *in situ* hybridization analysis and/or examination of ZEBrA patterns as: 1) showing differential expression in both RA and AId (RA and AId markers) and 2) showing differential expression in RA but not AId (RA unique markers). Both positive and negative markers were included in both sets to identify biological pathways under differential

regulation. The background consisted of the full set of genes present in the zebra finch Agilent microarray and previously used to define the RA transcriptome [24, 25, 81]; curation of this oligonucleotide microarray is described in Lovell et al. [87].

**Results**

To provide complete and precise definitions of the dorsal intermediate arcopallium (AId) and of RA, we mapped the expression of *SCN3B* in adult zebra finches using *in situ* hybridization. In males, the *SCN3B*-defined RA boundaries corresponded closely to cytoarchitectonic boundaries under Nissl (Fig. 2.1A,D). In contrast, the borders of *SCN3B*-defined AId could not be seen under Nissl, however this region contained neuronal cells with large somata that resemble the projection neurons found in RA, and thus may correspond to AId projection neurons [88, 89]. They contrast sharply with the smaller and densely packed cells in the adjacent ventral intermediate arcopallium (AIv; Fig. 2.1D), or the dorsal arcopallium (AD; not shown). The *SCN3B*-defined AId closely matches the region containing the projection terminals from the shell of the lateral magnocellular nucleus of the anterior nidopallium (LMAN) shell [83]. In females, where RA is atrophied [7, 42, 88] an *SCN3B*-defined AId, but not RA, was clearly visible (Fig. 2.1C). Importantly, both RA and AId were identifiable in adjacent sections via differential *PVALB in situ*, and RA was visible under Nissl staining as a small nucleus with high cell density directly medial to the *SCN3B*-defined AId (Fig. S1). SCN3B expression borders were then drawn on serial transverse *SCN3B* sections (200 µm intervals) throughout the whole extent of the arcopallium of adult birds. In males, *SCN3B*-defined AId occupied an extensive area, with a rostro-caudal extent (~0.1 P to 1.1 P) somewhat larger than RA (Fig. 2.1E left). No distinct

expression boundary was distinguishable between AId and RA at the core of RA (0.9 P in Fig. 2.1E left), in fact these two areas formed a medial-to-lateral continuum of low expression. At rostral or caudal levels RA and AId were separated by regions of high *SCN3B* expression (0.5 P or 1.1 P in Fig. 2.1E left). *SCN3B*-defined AId had a similar location in females (Fig. 2.1E right) but appeared smaller than in males caudally. In sagittal *SCN3B in situ* images (from ZEBrA), AId was distinguishable from the surrounding arcopallium as a core area of low expression lateral to RA (Fig. 2.1F) with distinct cytoarchitectonics and continuous with a rostral domain (AIr in Fig. 2.1F, middle) previously defined in Mello et al. [34].

To better characterize the molecular relationships between AId and RA, we next examined the *in situ* hybridization patterns of 162 genes that are differentially expressed in RA. This analysis included a set of 46 genes previously identified through microarray screenings as markers that RA in finches (and analogous nuclei in vocal learning birds) share with the laryngeal motor cortex (LMC) in humans [24] and 116 genes identified as RA markers in ZEBrA. For a set of 60 genes that included the 46 shared RA and LMC marker set and 14 RA markers from ZEBrA, we ran *in* situ hybridizations in frontal sections containing both RA and AId (examples in Fig. 2.2), noting that an assessment of the expression in AId had previously not been performed for most of these genes. For the other genes in the RA marker set, we evaluated expression in the sagittal image series available on the ZEBrA website. Among the 162 genes analyzed, numerous had much lower expression in RA and AId compared to the surrounding arcopallium (e.g. *SCN3B* and *SNCA* in Fig. 2.2), whereas other genes were more highly expressed in both nuclei compared to the surrounding arcopallium (e.g. *PVALB* and *LPL* in Fig. 2). Yet other genes were

only modestly differential in RA and AId (e.g. *SYNPR*, *GPM6A* and *RCAN2* in Fig. 2.2), showing regional expression level differences rather than highly differential patterns. Notably, some genes showed less differential expression compared to the surrounds in the dorso-medial part of AId close to the boundary with RA than in the more ventro-lateral AId. We refer to this dorso-medial area as the neck of AId (nAId; top left drawing in Fig. 2.2), and suggest that it may correspond to a transition zone between RA and AId as previously described based on connectivity [39]. Lastly, numerous other genes were only differentially expressed in RA but not in AId (e.g. KCNC1 and GABRE in Fig. 2.3). We note that the patterns for the set of genes assessed in frontal sections were qualitatively similar in both the left and right hemispheres.

Based on this analysis, we classified each gene as being either a unique marker of RA, which is more likely related to the neurobiology of learned vocalizations, or a marker of both RA and AId, which may represent features of motor control circuits rather than specializations unique to vocal-motor control (Table 1). Notably, of the shared markers of songbird RA and human LMC (n=46, Pfenning et al.), a large proportion (58%, n=25) were also differential in AId (Table 1, RA and AId markers), whereas only 42% (n=21) turned out to be RA unique markers (Table 1, RA unique markers). To provide quantitative support for our visual assessment, we calculated a ratio of expression levels in RA compared to AId for a subset of the genes examined. Genes that were classified as having qualitatively similar levels of expression in RA and AId (Table 1; e.g. *CNTNAP2* in Fig. 2.3A) were found to have ratios very close to 1 (Fig. 2.3B, black columns; SD: 0.036), whereas genes classified as unique positive or negative markers of RA (e.g., *KCNC1* and *GABRE*, or *SLIT1* in Fig. 3A, respectively) had expression ratios that were higher (>1) or lower

(<1) in RA than AId (Fig. 2.3B, green and blue columns, respectively). Plotting the densitometric ratios for all genes quantified revealed that both RA unique markers as well as RA and AId markers could be found over a wide range of expression levels. These data thus support our visual classification of *in situ* patterns based on relative differences in expression in RA vs AId.

We next examined whether all 162 RA markers were also differentially expressed in 12 additional molecularly defined arcopallial domains (collapsed from 19 in Mello et al. [34]). We found that >55% of RA markers were also differential in AId, followed distantly by other domains like AIr and AA (~22% and ~18%, respectively; Fig. 4). This suggests that RA and AId are more molecularly similar to each other than to other arcopallial domains. We followed up by asking whether the 98 markers of both RA and AId were also markers of other arcopallium domains. We found that >30% of RA/AId markers were also markers of AIr, followed distantly by AA, AMV, and AD (Fig. 2.5A). AIr is located rostral to and directly bordering AId, as best seen on sagittal sections (Fig. 2.5B, top left).

To further investigate the relationship between AId and RA, we performed a pathway enrichment analysis comparing sets of marker unique to RA with those that were both RA and AId markers, noting that a previous analysis [87] did not consider whether RA markers were also differential in AId. For both marker sets, significantly enriched annotations related to physiological features such as regulators of cell excitability (potassium channels, voltage gated channels), regulation of intracellular calcium levels or calcium related signaling, and neuronal connectivity (summarized in Table 2). While RA and AId seem to share most of their specialized

molecular pathways, the sets of markers uniquely expressed in RA differentiate it from AId. We suggest that this set of genes that are unique to RA are more likely to contribute to the unique properties of RA and its role in the neurobiology of learned vocalizations (Table 1, RA unique markers).

RA in zebra finches undergoes marked developmental changes in morphology, connectivity, physiology and gene expression [88, 90-94], but except for tract-tracing data [95], little is known with regards to age differences in AId. We therefore asked whether adult AId markers also define AId in 20 dph juvenile males entering the sensory phase of vocal learning when they can start to form an auditory memory of the tutor song. They are also pre-vocal, as this age is prior to the formation of the HVC-to-RA projection, which marks the beginning of singing and of the babbling phase [88, 96]. RA has also not started its massive expansion in males or regression in females [88]. *SNCA*, a robust differential marker of adult AId, was highly differential in 20 dph AId compared to surrounds, and even though RA is much smaller at this age (confirmed under Nissl), it formed a continuum of low *SNCA* expression with AId (Fig. 2.6A-B), noting that expression within AId was restricted to sparse cells as in adults (Fig. 2.6C, left). The positive marker *PVALB* showed similarly high expression in juvenile as in adult AId (Fig. 2.6C, middle). Thus, AId is already present in juveniles and expresses some molecular features of adult AId. In contrast, *SCN3B* showed considerable expression in juvenile AId (Fig. 2.6C, right) and was thus less differential compared to the adjacent arcopallium than in adults. This suggests that AId is not fully mature and undergoes further molecular differentiation until the birds reach adulthood.

Previously, immediate early gene expression elicited from movement has been described in an AId-like area in vocal learning birds (songbirds, parrots, hummingbirds) and in a possibly related part of the arcopallium in a non-vocal learning avian species (doves) [35]. To further investigate if AId is present in presumed avian non-vocal learners, we next asked if molecular markers of adult AId in zebra finch also define an AId-like area in suboscine species. Suboscines, the sister taxa to the oscines (i.e. songbirds), are also passerines (perching birds) and similar to songbirds in terms of anatomy and physiology, but are generally considered to lack RA and other telencephalic vocal nuclei [80, 97]. Differential expression of *PVALB* and androgen receptors can be interpreted as suggestive evidence of an AId-like area in the suboscine families Tyrannidae [97] and Pipridae [98], respectively. Here we examined males from representative species from the Thamnophilidae (Willis's antbird; *Cercomacroides laeta*) and Dendrocolaptinae (Straight-billed woodcreeper; *Dendroplex picus*) families. *In situ* hybridization for the robust RA and AId marker *SNCA* in finches showed marked downregulation in a very similar area as finch AId in both suboscine species (Fig. 2.7), noting that an RA-like nucleus could not be identified in either species with *SNCA* or by Nissl.

**Discussion**

We investigated the expression of a large set of differential arcopallial markers in zebra finches in order to better define AId in a songbird, examine the molecular relationships between RA and AId and other arcopallial domains, and more precisely identify molecular features unique to the vocal motor system. Our results provide a clear delineation of AId boundaries in adults of

both sexes, and support a closer similarity between RA and AId compared to other arcopallial domains. We identify molecular specializations that are common to both RA and AId and possibly related to diverse aspects of motor control, as well as those that are unique to RA and more likely associated with vocal-motor control. We also show AId is likely present in birds that do not learn their vocalizations (non-singing females, pre-vocal juvenile males, non-vocal learner suboscines), consistent with a broader role of AId in motor control and a possible evolutionary and developmental origin of RA as a vocal-motor specialization of AId.

AId has been previously defined as a subregion of the songbird intermediate arcopallium that has distinct connectivity (discussed below) but whose boundaries are not readily identifiable under Nissl staining. Previous studies have shown that AId has distinct molecular features, including prominent expression of *PVALB* and a lack of expression of *SCN3B* [34, 81, 99]. Here we have more clearly delineated AId boundaries, and shown that it extends over a large portion of the intermediate arcopallium in both sexes. We note that given its shape and size, AId would be a difficult region to fully cover with stereotaxic injections, thus studies of connectivity need to use multiple injections over a range of coordinates to ensure full targeting, or to precisely track the injection position to address possible topography. Based on molecular similarity and spatial proximity, we also suggest that AIr might be a specialized rostral expansion of AId with a yet to be determined function.

RA and AId have several anatomical features in common, suggesting some similar functions and a close evolutionary relationship. Whereas RA receives input from HVC and from the LMAN

core, the latter projection connecting the anterior forebrain and vocal-motor pathways, AId receives input from the LMAN shell and from the dorsal caudolateral nidopallium (dNCL) lateral to HVC[82, 83, 95, 100]. The projection from dNCL to AId appears to be topographic [11, 83, 95, 100] suggesting that a topographic organization might also be present in a medial to lateral map in AId, with possible somatotopy. Furthermore, both RA and AId are part of the intermediate arcopallium, which originates descending somatic projections [36] and is considered part of a general motor pathway [71]. RA projects to the medullary vocal-motor nucleus nXIIts [74] and AId projects to targets in midbrain, pons, medulla, and possibly spinal cord, projections that could be considered analogous to the corticobulbar and corticospinal tracts in mammals [83]. Accordingly, RA and AId as well as other parts of the arcopallium (e.g. dorsal, AD) show enriched expression of markers of deep layers of the mammalian cortex [34, 37] where long descending projections originate, consistent with the idea that AId is analogous to, or contains an avian analog of deep layers of mammalian cortex [40]. Interestingly, examination of the Allen Brain Atlas mouse brain *in situ* hybridization data [101] shows that some AId markers exhibit an enrichment in deep layers of the mouse motor cortex whereas others are broad deep layer cortical markers (Fig. S2).

RA and AId also seem to have analogous roles in motor control. For RA, there is evidence of severe vocal deficits after lesioning [11] and evidence of activation during singing based on electrophysiological recordings [18, 73] and immediate early gene expression [84]. Although less studied, evidence for a motor control function for AId comes from the immediate early gene activation during movements such as wing flapping, hopping, and pecking [35, 75], and the severe motor deficits associated with large lesions [39]. Interestingly, the age differences

we observed in gene expression suggest that AId undergoes molecular changes during development, which could be associated with vocal or other motor learning refinement. AId has been hypothesized to be directly involved in vocal learning [85]. We note, though, that the adjacent ventral intermediate arcopallium (AIv) is an auditory area that responds to song playbacks [102] and may play a role in vocal learning [103], thus the possible separate roles of these adjacent areas in auditory processing, vocal learning, and motor function remain to be conclusively determined.

RA and AId have previously been shown to share a few molecular markers in adult males but here we considerably expand that evidence, and show that AId is the arcopallial domain that shares most known markers with RA. Importantly, we have found that 25 out of the 46 genes previously identified as shared markers of songbird RA and human laryngeal motor cortex (LMC) are in fact markers of both RA and AId (Table 1). This finding argues that over half of the shared RA/LMC markers may not be uniquely associated with the vocal motor pathway or vocal control, but could perhaps subserve a broader array of somatic motor control functions. Among identified enriched pathways for these RA and AId markers is the rapid depolarization pathway containing sodium and calcium channel genes known to be differentially regulated in the song system [104], and the axon guidance pathway containing *GAP43*, which is a shared marker of RA and LMC [105]. Particularly noticeable was the very low expression in RA and AId of *SNCA*, previously shown to be transcriptionally regulated in LMAN during song learning but constitutively downregulated in HVC and RA [106]. *SNCA* encodes α-synuclein, a major component of the pathological Lewy body aggregates associated with Parkinson's disease, Lewy bodies dementia, multiple systems atrophy, and a subset of Alzheimer's disease cases [107-

109]. Aberrant transcriptional activation of *SNCA* during aging may contribute to the motor deficits seen in patients with Lewy body pathologies [110]. We suggest that downregulated *SNCA* in both vocal and presumed adjacent motor areas may serve as a protective mechanism for maintaining motor control circuits during aging in birds. Notably, we also show that a much larger set of RA markers not previously described as shared markers with LMC are also AId markers (Table 1). This includes other genes related to axonal guidance like *PLXNA1* and *ROBO1*, suggesting broader roles in motor connectivity rather than a specific role in vocal-motor circuits, as previously concluded for the latter [99]. It would be interesting to ascertain in future studies whether RA and AId markers are also differential in human primary motor cortex, using more refined dissections than those used for the human dataset in Pfenning et al [24].

Importantly, we have also identified molecular specializations that are unique to RA rather than common to RA and AId. These genes, which include 21 shared RA and LMC markers from Pfenning et al. [24], now represent a better validated set of molecular features unique to this key nucleus in the circuitry for learned vocalizations. Intriguingly, many of the enriched pathways in the set of RA unique markers are similar to those of the RA and AId markers, even though the two gene lists are distinct, suggesting that specific genes within a given family may confer unique properties to each area. Axon guidance, for example, seems to be of particular relevance to the neurobiology of learned vocalizations, and axon guidance pathways have been shown to be targets for human speech and language disorders [111]. Precise vocal production relies on direct cortical projections to brainstem motor neurons that control the vocal organ, exemplified by the projection of RA to nXIIts in songbirds. Non-vocal learning animals do not

have this direct projection and instead are thought to have only indirect projections from cortical motor areas to the vocal hindbrain via the midbrain [112]. A unique set of axon guidance cues likely enable the vocal-related projection from RA to the brainstem vocal nuclei to be formed and/or maintained. Some examples include *SLIT1* and *PLXNC1*, both shared markers of RA and LMC that we showed here not to be differential in finch AId, in agreement with previous observations [99]. *SLIT1* is a target of FOXP2, a transcription factor linked to speech developmental disorder [113], and it is differentially regulated during vocal development [99], possibly contributing to the establishment of the RA to brainstem connection. *PLXNC1* has undergone a partial duplication unique to parrots, and has been suggested as potentially linked to the expanded vocal and imitative abilities of that taxon [114]. Other axon guidance genes shown here to be differentially expressed in RA but not AId include *RELN* and *PLNXA4*, both of which also contribute to neuronal migration [115, 116], and *RTN4R* and *LINGO1*, both key components of the Nogo receptor complex and known regulators of axonal growth and myelination [116, 117].

Another class of proteins likely of functional relevance to RA is potassium channels, which are major determinants of neuronal excitability. Potassium channels are broadly expressed throughout the brain but song nuclei exhibit unique expression profiles of both potassium channels [118] and other ion channel families. These observations suggest that vocal nuclei are sites of differential regulation of intrinsic excitability, consistent with growing evidence that modulation of intrinsic excitable features may play important roles in regulating properties of the vocal learning circuitry [119, 120]. A potassium channel gene uniquely enriched in RA is the voltage gated potassium channel subunit *KCNC1*, which encodes the $K_V3.1$ protein. *KCNC1* has

been associated with high-frequency firing in auditory brainstem neurons [121], and an upregulation in RA likely contributes to the high-frequency firing capability of RA neurons [72]. We also obtained evidence that genes associated with neurotransmission and synaptic function are uniquely differential in RA, including several neurotransmitter/neuromodulatory receptors (*GABRB3*, *GABRE*, *GRK3*, *CHRM4*, *HTR1B*), in contrast to related genes also differential in AId and likely more related to synaptic regulation in the broader context of motor function (*GRIN2B*, *GRM3*, *HTR2A*). It also worth noting that some RA unique markers are transcription factors that potentially exert marked but still unexplored roles in regulating the differentiation and function of vocal circuits. This includes *SAP30*, which as part of a large histone deacetylation complex can regulate transcription and chromatin remodeling [122, 123], *NEUROD6*, which interacts with several other factors (*TBR1, FEZF2, FOXG1, SATB2*, *EMX1*) linked to cortical development [124] and is involved in regulation of callosal projections [125], and *RORA*, which has been implicated in cortical and cerebellar development [126] and autism [127].

We note that RA in both juvenile males and adult females, while smaller than in adult males, is still continuous with the medial end of AId. It thus appears that as RA undergoes its developmental growth in males, it likely expands medially and ventrally so that in adults it ends closely adjacent to the medial arcopallium (AM) and to the RA cup, the latter considered part of AIv and related to auditory processing [68, 70, 128]. However, RA is very distinct from AM and AIv, both molecularly and in terms of connectivity [70, 129]. Our observations support a much closer relationship between RA and AId. It is not known, however, if AId has unique molecular features, as all known markers of AId are genes initially identified as RA markers. Furthermore,

the evidence for molecular similarities between AId in songbirds and in non-vocal learning suboscines is consistent with a broader motor function for AId, as contrasted to the exceptional specialization of RA for vocal-motor function. It also supports the notion that RA may have evolved as a specialization of a primordial motor region present in birds independently of the occurrence of learned vocalizations, referred to as the motor theory for vocal learning origin [35], favoring it over an auditory origin. It is important to note that under this hypothesis, RA and AId would not have to be the sole motor output of the zebra finch arcopallium. The AD, located directly dorsal to RA and AId also has somatic-like projections to the thalamus, midbrain, and brainstem [36] and may also represent an avian analog to layers 5/6 of motor cortical areas. While several arcopallial domains express markers of both mammalian cortical and amygdalar subdivisions, AD predominantly expresses markers of layer 6 cortical neurons [34]. It is unknown if neurons analogous to pyramidal projection neurons in layers 5/6 are intermixed throughout the arcopallium or segregated into separate domains.

In summary, we provide molecular evidence for a close relationship between RA and AId, as well as clearly identify molecular specializations unique to RA. Our findings are consistent with AId being an ancestral motor region from which the vocal nucleus RA may have evolved. The data also provide an invaluable source of candidate genes for future studies on specialized vocal learning mechanisms.

| | RA unique markers | RA and AId markers |
|---|---|---|
| Shared human LMC markers | C1QL3, CNTN3, CYGB, DAAM1, DGKZ, DPYSL3, GABRB3, GPRC5B, KCTD15, LINGO1, MTCL1, NEUROD6, NEXMIF, NTRK2, RORA, RTN4R, SAP30, SLIT1, SMAP1, UCHL1, YWHAH | B3GAT1, CDH4, CDH11, DOCK4, ETNK1, FAM49A, GAP43, GNG2, GPM6A, LYPD1, NECAB2, NOL4, PCDH17, PLXNC1, PPFIA2, PPP2R5C, PVALB, SLC25A22, SNCA, SYNPR, SYT17, TENM3, TMEFF2, VIP, ZBTB18 |
| Other markers | ABCG4, ACKR3, ADCYAP1, ARPC5, B3GNT2, BTBD10, CAMK2A, CAMKK1, CDH8, CHRM4, CRHR2, CTSL, DPP6, GABRE, GRK3, HTR1B, IGF1, IL1RAPL2, KCNA6, KCNC1, KCNF1, KCNJ6, KCNK12, KCNS2, KCNT2, KCTD20, KIF26B, LRP8, MAP4, MGAT4C, NTNG1, P4HA2, PLXNA4, PTPRN2, PTPRU, RELN, RHOB, RIMS4, SCN2A, SEPT12, SEPT6, SH3BP5, SLC8A1, STK26, STMN1, SYBU, TNFAIP8L3 | ADAM23, ADRA1D, ANXA6, APOH, ARHGDIB, ATP2A3, ATP2B2, BAIAP2, CABP1, CACNA1E, CACNA2D2, CACNB2, CAMTA1, CD99L2, CDH9, CNTN4, CNTNAP2, CPNE2, CRHR1, DHCR24, FAM163B, FLRT2, GDA, GLRA2, GLRA4, GRIN2B, GRM3, HTR2A, HRH3L, JPT1, KCNAB1, KCNAB2, KCNS1, KCTD12, KCTD16, KCTD3, LPL, MPZL1, , NEFL, NEFM, NEGR1, NRXN1, NTSR1, PAK6, PCP4, PLD1, PLS3, PLXNA1, PRKAR1B, PTPRZ1, QSOX1, RCAN2, RGS4, ROBO1, SAP30L, SCN1A, SCN3B, SCN4B, SCN8A, SCUBE1, SEMA7A, SLC24A2, SLC4A4, SLC6A7, SYF2, SYNGR3, UGT8, UNC5D, WASF1 |

Table 2.1: Genes differentially expressed in RA only (RA unique markers) are separated from genes that are differential markers of both areas (RA and AId markers). For both groups, we further indicate whether RA markers were previously reported or not as shared markers with human LMC.

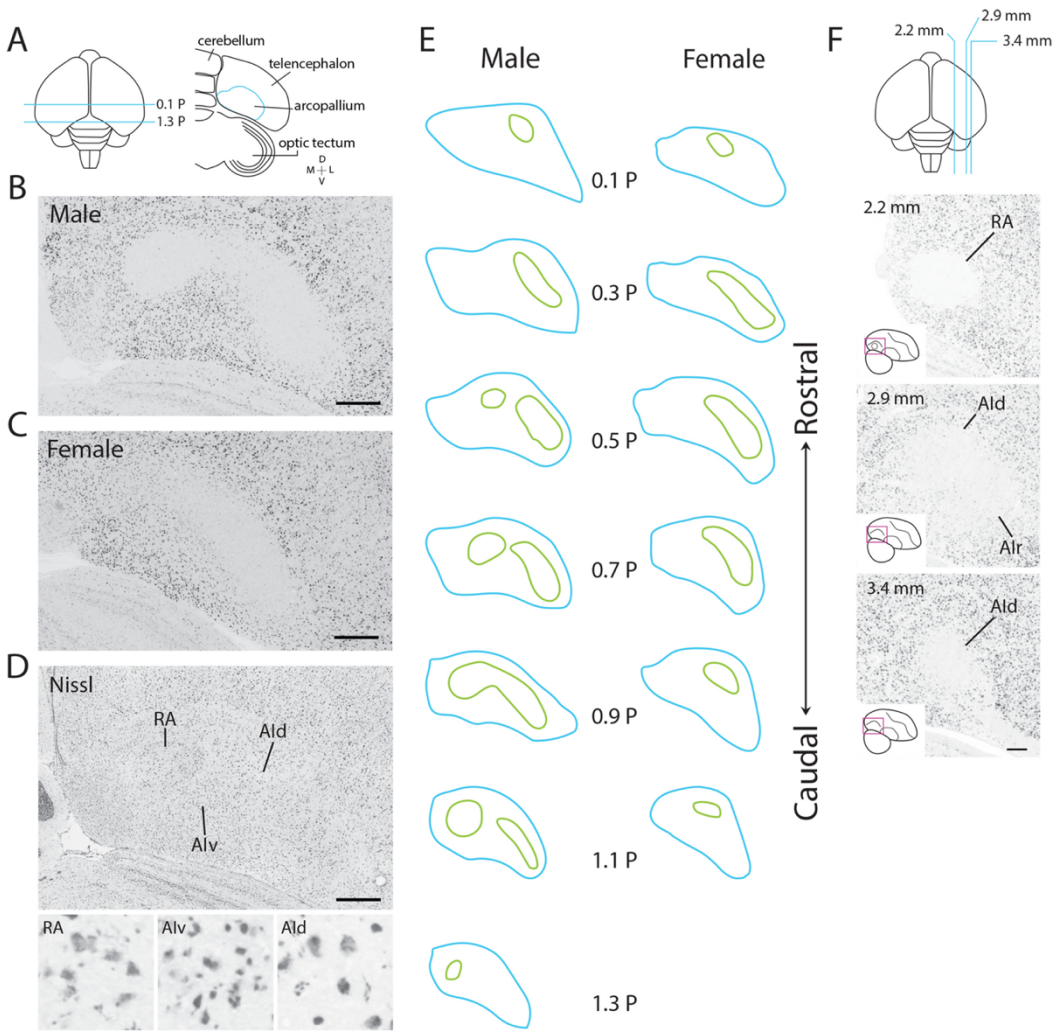



Figure 2.1: Molecular definition of RA and AId in adult zebra finches. (A) Top left: Top-down view of a schematic drawing of the zebra finch brain; blue lines indicate the range of frontal sections examined in this study. Top right: Drawing of a frontal section at 0.7P; blue line indicates the boundaries of the arcopallium, seen under Nissl staining. (B-C) *SCN3B in situ* hybridization images from a male and a female. RA and AId appear continuous in the male, and RA is indistinguishable in the female. (D) Nissl-stained frontal section through arcopallium at the center of RA in a male; RA, but not AId, has clear cytoarchitectonic boundaries. Small panels show high power views (100 x 100 µm images) within RA, AIv, and AId. (E) Drawings depicting *SCN3B* expression boundaries (green) in serial frontal sections through the arcopallium (blue) of adult male (left) and female (right) zebra finches. (F) Sagittal series of *SCN3B in situ* hybridization images through the arcopallium, reproduced from ZEBrA (www.zebrafinchatlas.org); section level is indicated by the blue lines in the schematic drawing at the top. Abbreviations: AId: dorsal intermediate arcopallium, AIr: rostral intermediate arcopallium, AIv: ventral intermediate arcopallium, RA: robust nucleus of the arcopallium. Scale bar: 400 µm for all images.

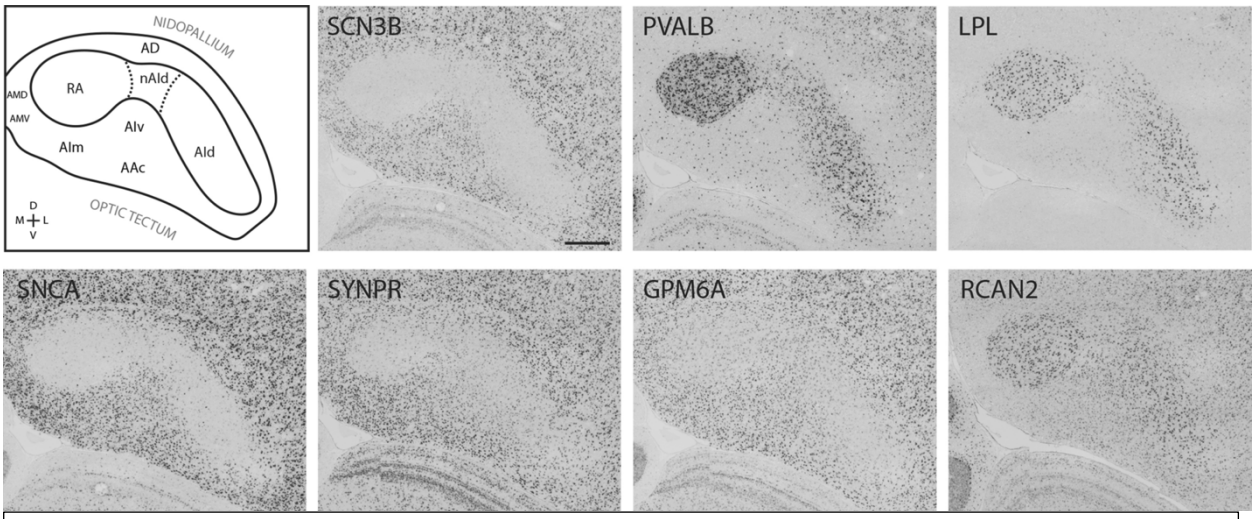



Figure 2.2: Expression patterns of RA and AId markers. Top left: Drawing of the zebra finch arcopallium in the frontal plane, depicting structures shown in all other panels. RA and AId were defined based on the SCN3B expression pattern in the next panel, placement of other domains derives from Fig 17 in Mello et al.[44]. Other panels: *in situ* hybridization images for various RA and AId markers. Scale bar: 400 µm for all images. Abbreviations: AAc: caudal anterior arcopallium, AD: dorsal arcopallium, AId: dorsal intermediate arcopallium, AIm: medial intermediate arcopallium, AIv: ventral intermediate arcopallium, AMD: dorsal medial arcopallium, AMV: ventral medial arcopallium, nAId: neck of the dorsal intermediate arcopallium, RA: robust nucleus of the arcopallium.

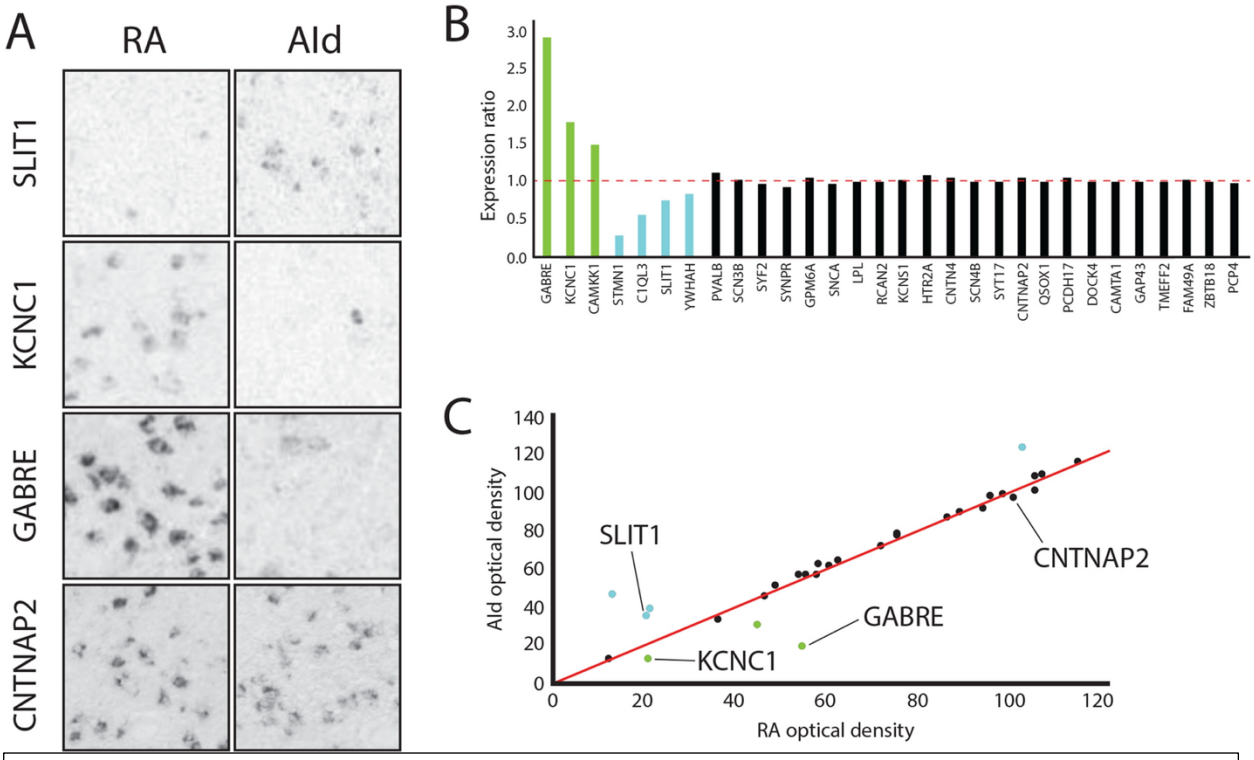



Figure 2.3: Defining molecular specializations unique to RA or common to both RA and AId. (A) High magnification (200x200 µm) *in situ* hybridization images of RA and AId showing cell-level expression of select genes RA unique and RA and AId markers. *SLIT1* (top) and *KCNC1* and *GABRE* (middle) are respectively negative and positive markers unique to RA, whereas expression of *CNTNAP2* (bottom) is similar in RA and AId. (B) Expression ratio (optical density within RA/optical density within AId) for genes visually determined to be positive (green) or negative (blue) markers of RA only, or markers of both RA and AId (black). A ratio of 1 (dashed red line) corresponds to a gene equally expressed in RA and AId, and ratios >1 or <1 correspond, respectively, to positive or negative RA unique markers. (C) A scatterplot of expression ratio values for the genes in (B), with RA unique markers deviating from the 1:1 expression ratio line (red).

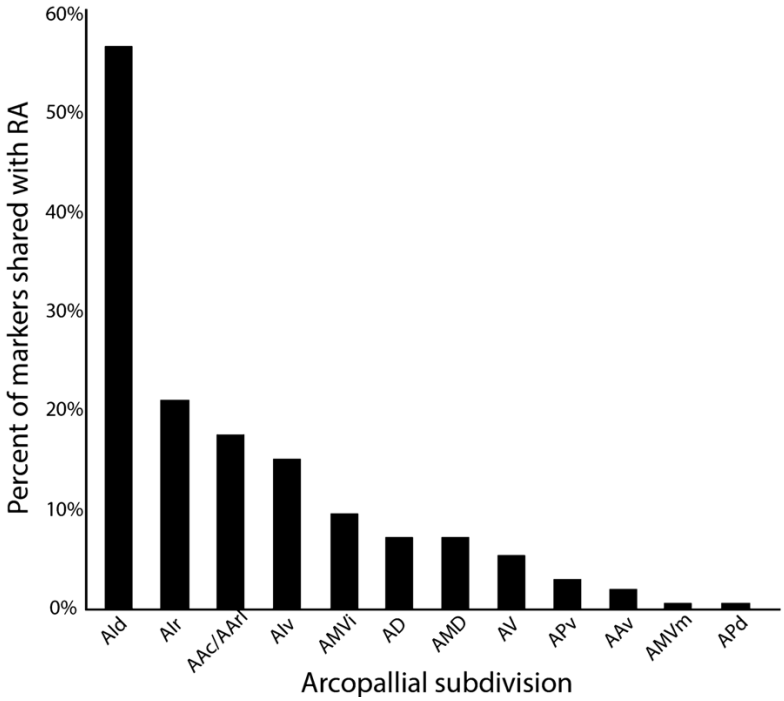



Figure 2.4: Molecular relationship between RA and other arcopallium domains. The arcopallial expression patterns of 162 RA markers were analyzed based on *in situ* hybridization data from the present study and from ZEBrA. Plotted are the percentages of RA markers that were also considered markers of other arcopallial domains; individual genes can be represented in multiple columns. Abbreviations: For a complete list of abbreviations see the legend in Figure 5.

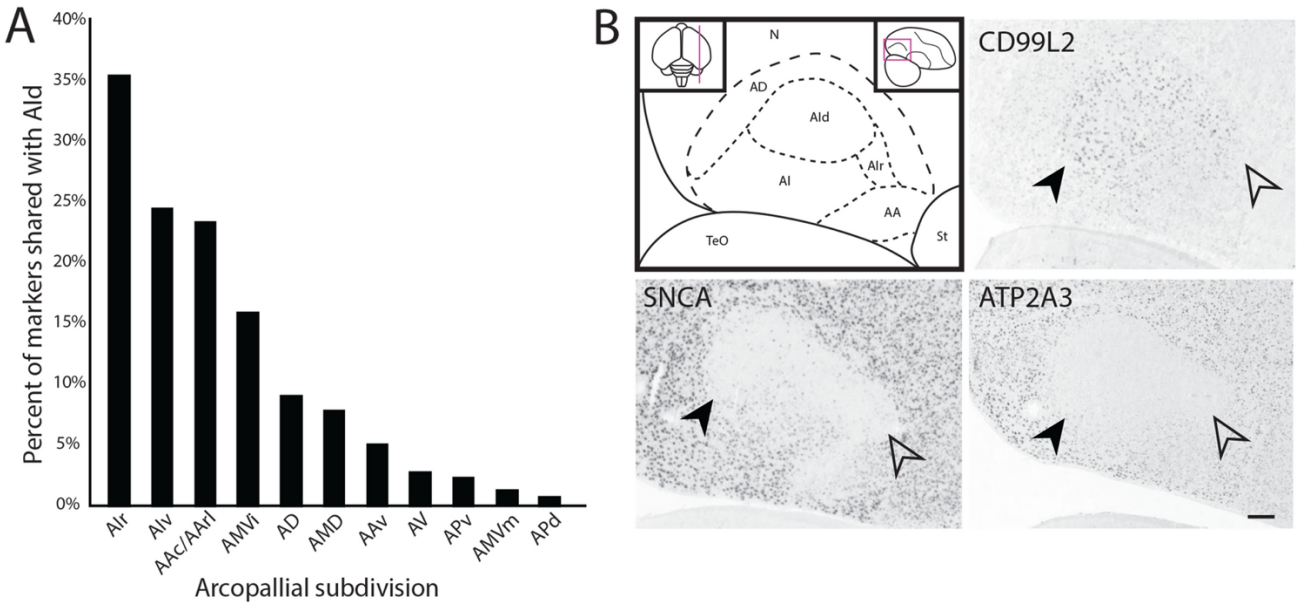



Figure 2.5: Relationship between AId and other arcopallial subdivisions. (A) The arcopallial expression patterns of 98 RA and AId markers were analyzed based on *in situ* hybridization data from ZEBrA. Plotted are the percentages of RA markers that were also considered markers of other arcopallial subdomains; individual genes can be represented in multiple columns. (B) Top left: drawing of arcopallium and its main subdomains on a sagittal section; top left inset indicates the position of the section (~2.9 mm from the midline) on a top-down view of the brain; red rectangle in the top right inset indicates the area shown in the main drawing and other panels. Other panels: *In situ* hybridization images of positive (*CD99L2*), negative (*ATP2A3*), and sparse cell (*SNCA*) markers of AId (black arrowheads) and AIr (empty arrowheads). Scale bar: 400 µm. Abbreviations: AA: anterior arcopallium, AAc: caudal part of the anterior arcopallium, AArl: rostro-lateral part of the anterior arcopallium, AAv: ventral part of the anterior arcopallium, AD: dorsal arcopallium, AId: dorsal intermediate arcopallium, AIr: rostral intermediate arcopallium, AIv: ventral intermediate arcopallium, AMVi: intermediate part of the medial ventral arcopallium, AMD: medial dorsal arcopallium, AMVm: medial part of the medial ventral arcopallium, APv: ventral part of the posterior arcopallium, APd: dorsal part of the posterior arcopallium, AV: ventral arcopallium, N: nidopallium, St: striatum, TeO: optic tectum.

| Pathway name | RA unique markers | RA and AId markers |
|---|---|---|
| Potassium channels | KCNA6, KCNC1, KCNF1, KCNJ6, KCNS2 | KCNAB1, KCNAB2, KCNS1 |
| Axon guidance | ARPC5, DPYSL3, PLXNA4, RELN, RHOB, SCN2A, SLIT1 | CACNB2, GAP43, GRIN2B, SEMA7A, SCN1A, SCN3B, SCN4B, SCN8A, PAK6, PLXNA1, PLXNC1, ROBO1, UNC5D |
| GPCR binding and signaling | ACKR3, ADCYAP1, CHRM4, CRHR2, DGKZ, GRK3, HTR1B, LRP8, RHOB | ADRA1D,HRT2A, CACNB2, CRHR1, GNG2, GRIN2B, GRM3, LPL, NTSR1, PRKAR1B, RGS4, VIP |

Table 2.2: Summary of enriched pathways and related genes for RA unique and RA and AId marker sets.

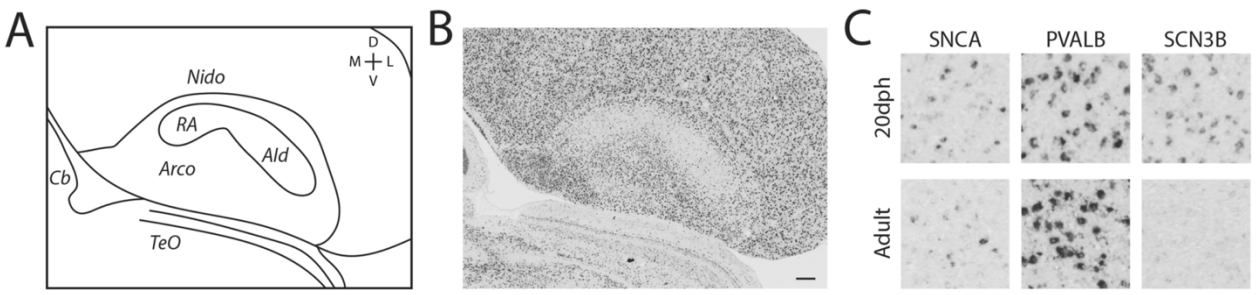



Figure 2.6: Defining AId in juvenile male zebra finch. (A) Drawing of the arcopallium in frontal section through the core of RA in a 20 dph male zebra finch (based on B), depicting the continuous area of low *SNCA* expression with sparse labeled cells that includes both RA and AId. (B) *In situ* hybridization image of *SNCA* in a 20 dph male zebra finch. C) High magnification (200x200 µm) *in situ* hybridization images of AId for select adult RA and AId markers, comparing cell level expression in 20 dph juvenile and adult males. Scale bar: 200 µm. Abbreviations: Arco: arcopallium, AId: dorsal intermediate arcopallium, Cb: cerebellum, Nido: nidopallium, RA: robust nucleus of the arcopallium, TeO: optic tectum.

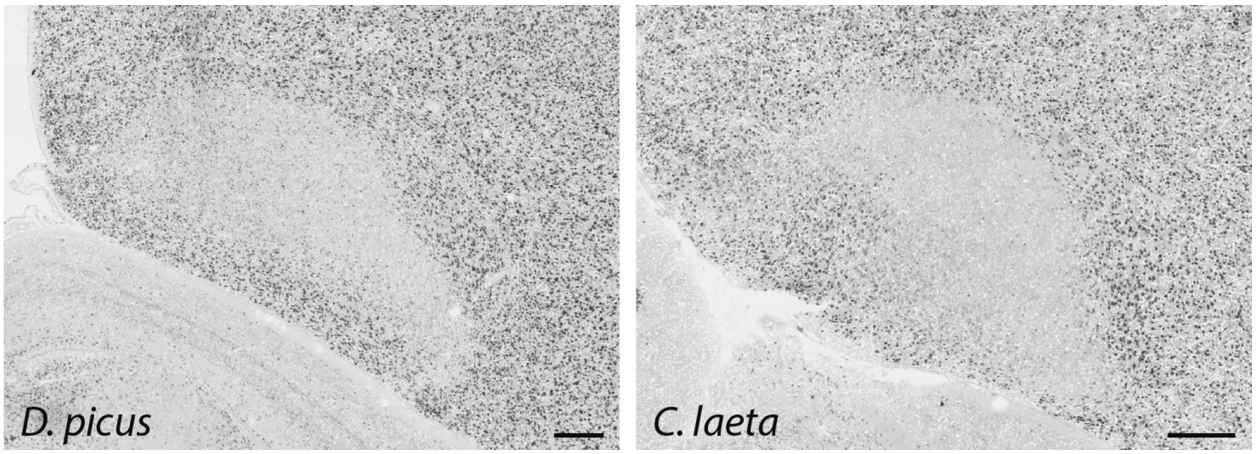
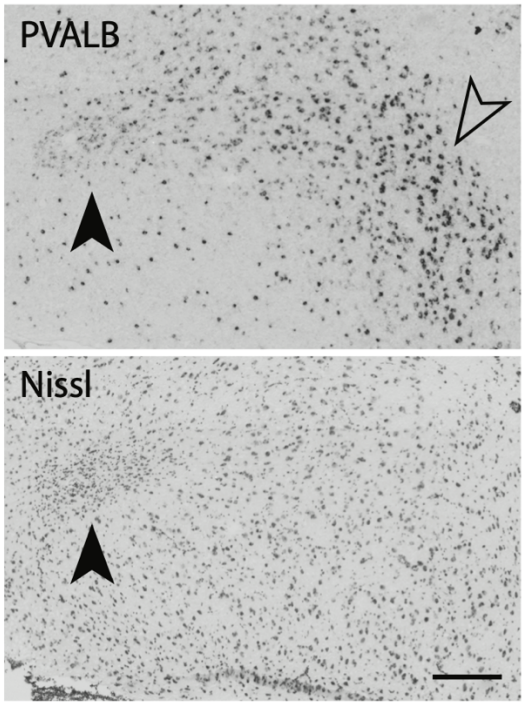
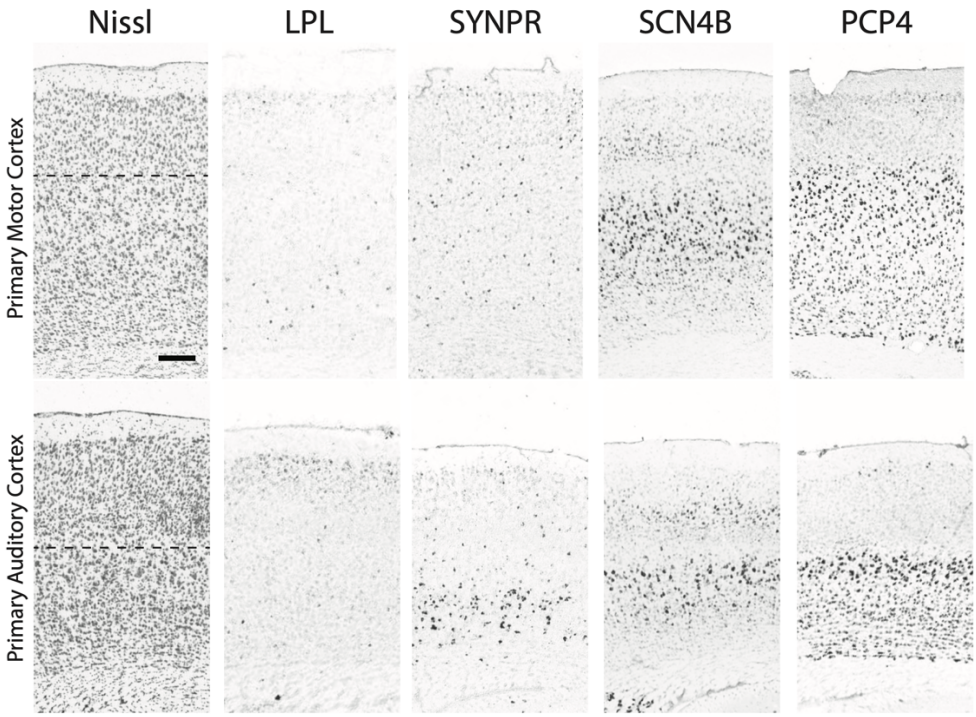



Figure 2.7: Defining AId in a suboscine. Representative *in situ* hybridization images of frontal sections through the arcopallium in two sub-oscine species, processed for an RA and AId marker (*SNCA*) from adult male zebra finches. Scale bar: 400 µm.




Supplemental Figure 1: Identification of RA in adult females. High power view of adjacent frontal arcopallial sections processed for *PVALB* (top) or stained for Nissl (bottom). Female RA can be identified by differential *PVALB* expression or higher cell density (closed arrow) compared to adjacent areas, but not by SCN3B expression (see Fig. 1C; *SCN3B* image from section adjacent to those shown here). *PVALB* is a marker of AId in the adult female finch (open arrow). Scale bar: 200 µm.




Supplemental Figure 2: Select RA and AId markers are also markers of deep cortical layers in mouse. *In situ* hybridization images from the Allen Institute Mouse Brain Atlas[78] for *LPL* (positive) and SYNPR (negative) in deep layers of motor cortex compared to auditory cortex. *SCN4B* and *PCP4* (both positive markers of RA/AId) are positive deep layer markers in both motor and auditory cortex. Dashed line in Nissl separates supragranular from infragranular layers. Scale bar: 200 µm for all images.

# Chapter 3

## Molecular specializations in a vocal-motor brain region at the cell type level

Alexander A. Nevue, Benjamin M. Zemel, Samantha R. Friedrich, Henrique von Gersdorff, Claudio V. Mello. AAN and CVM conceived the study. AAN performed all molecular experiments and analysis. BMZ performed and analyzed electrophysiology experiments. SRF helped develop interactive applications. HvG supervised the electrophysiology experiments. AAN and CVM wrote the paper.

**Abstract**

Understanding how novel brain circuits arise in evolution to support lineage specific behaviors requires defining the cellular and molecular specializations of those circuits, and the neuroanatomical context within which they arise. Songbirds learn their vocalizations from a tutor via imitation and possess unique forebrain circuitry that supports this complex rare behavior. The robust nucleus of the arcopallium (RA) controls the acoustic features of learned vocalizations and is analogous to deep layer pyramidal neurons of the human laryngeal motor

cortex. Using single nuclei transcriptomics and spatial gene expression mapping, we have identified cell types and molecular specializations present in RA but absent in adjacent cortical-like somatic motor and auditory areas. This includes an RA-specific vocal projection neuron, differential usage of inhibitory neuron subtypes, and unique glial specializations. The data are presented using interactive apps that facilitate integration of cell level molecular data with our spatial gene expression brain atlas ZEBrA. Our findings define molecular specializations unique to RA that likely reflect adaptations key to the physiology and evolution of learned vocalization circuits.

## Introduction

Understanding how complex behaviors are subserved by specialized circuits requires a thorough understanding of the distinct cell types and connections that constitute those circuits, as well as the neuroanatomical context in which they arose in evolution [130, 131]. To further understand these fundamental questions, it is key to investigate an organism with a well characterized behavior and associated brain circuitry, so that cellular specializations within behaviorally relevant pathways can be defined. Importantly, the identification of true molecular specializations of dedicated circuits requires careful contrasting analysis that takes into account neuroanatomical context and the broader areas from which these circuits evolved.

Vocal control circuits in vocal learning birds provide one of the most robust examples of specialized brain circuits that support lineage-specific behaviors [40]. In songbirds, parrots and hummingbirds, a dedicated circuit consisting of a set of interconnected forebrain nuclei enables

the acquisition and production of learned vocalizations. Notably, the architecture of avian vocal circuits reflect the general nuclear organization of the avian brain, in contrast to the cortical layered organization in mammals, yet they retain important connectivity features of cortical microcircuits [12]. Among the extensively characterized vocal nuclei in songbirds, the robust nucleus of the arcopallium (RA) is of particular interest because it provides the sole output of the forebrain vocal circuit and plays key roles in modulating the acoustic features of song [18]. RA is located in the intermediate arcopallium (AI), an avian brain structure with descending somatic projections and considered analogous to deep layers of mammalian motor cortex [132] [34]. The direct projections from RA to brainstem areas that control vocalizations and breathing [11] are similar to those of the human laryngeal motor cortex [33, 40]. RA neurons are highly specialized and capable of firing at very high frequencies [18, 91] due to unique ion channel specializations [45]. Despite its distinct physiological features and critical roles in producing song, some of the most basic features of RA such as cell type composition have yet to be fully characterized.

Previous efforts seeking to identify unique molecular markers of RA by differential screenings [24, 25] discovered molecular features of song nuclei that may support song behavior, as well as provided insights into their evolution. This includes the finding of convergently evolved differential markers shared by RA in songbirds and laryngeal motor cortex in humans, despite the independent evolution of these two lineages [24]. However, because the differential screening in Pfenning et al. (2014) contrasted RA with ventral AI (AIv), an avian auditory area, many of the molecular specializations proposed as unique to RA were later also found in the

adjacent dorsal AI (AId) of zebra finches [38], and thus are not specific to RA. The recent use of cellular transcriptomics partially addressed the need to define distinct cell types within vocal nuclei [133], to gain insight into how genes and molecular pathways operate within the context of specific cell populations. That study provided evidence suggesting distinct developmental origins for parts of the avian pallium and mammalian cortex despite a high degree of conservation of markers. However, it also did not address the neuroanatomical context, thus it is unclear if the cell types described are actually unique to vocal nuclei or reflect broader features of the corresponding brain areas.

In order to better understand the specialized functions and possible origins of RA, it is critical to define the molecular and cellular architecture of RA in relation to AId and the rest of the AI. AId is considered a motor cortical analog involved in somatic motor function [35, 39, 75] and broadly present in both vocal learner and non-learner birds (Nevue et al, 2020). Its close proximity and similar involvement in motor function as RA led to the hypothesis that RA may have evolved in songbirds as a differentiation or specialization of a more ancestral AId [35, 40], a notion supported by shared gene expression profiles between RA and AId, and molecular similarities in AId between songbirds and sub-oscines [38]. The more ventral AIv, in contrast, while also considered broadly present in birds, is thought to be involved in auditory processing [39]and likely expresses different markers than AId. A close comparison of RA with both AId and AIv arguably offers the best opportunity to identify cellular and molecular features truly unique to RA.

Here we performed single nucleus RNA sequencing (snRNAseq) on the intermediate arcopallium, including RA, AId, and AIv, combined with *in situ* hybridization to assess the spatial distribution of the identified cell types. We uncovered a multitude of molecular and cellular specializations unique to RA, including evidence supporting a specialized type of excitatory neuron, specialized features of some GABAergic subtypes, and non-neuronal cell specializations. Importantly, these features are absent in the adjacent AI subdivisions. These findings provide new insights into the molecular organization, physiology and evolution of a key area that enables the production of learned vocalizations. We have also developed interactive apps that allow integration of the present cell type profiling and developmental transcriptome datasets (Friedrich et al, 2022) with the spatial mapping of *in situ* hybridization data in the ZEBrA atlas [22]. These integrated resources will greatly facilitate data mining and exploring further the molecular genetic basis of learned vocal behavior, as well as the evolution of vocal control circuits.

## Results

*Cell type composition of the intermediate arcopallium (AI)*

To determine the cell type composition of the AI in adult male zebra finches, we microdissected an area encompassing RA, as well as the dorsal and ventral AI (AId and AIv), analogous to deep layer motor and sensory (auditory) cortices [12, 34], respectively (Supplemental Fig. 1A) and isolated nuclei for snRNA-seq [134] (Fig. 3.1A). We identified 13 highly resolved clusters

representing 8 classes of neurons, 2 classes of astrocytes (Astro_1-2), and one cluster each of oligodendrocytes, microglia, and endothelial cells (Fig. 3.1B, C, Supplemental Fig. 1B), based on the expression of established cell type markers [135, 136]. Among neuronal clusters, 4 were identified as excitatory (Ex_1-4) and 4 as inhibitory (In_1-4). There was a near 50/50 ratio in the numbers of neuronal vs. non-neuronal cells, with astrocytes being the most prevalent class and excitatory neurons more prevalent than inhibitory neurons (Fig. 3.1D). Astrocytes and oligodendrocytes representing a majority of non-neuronal cells is consistent with estimations from human cortex [137].

Because these cell clusters were derived from tissue containing multiple AI subregions and spatial information is lost in snRNA-seq, we next performed *in situ* hybridization for cell class proxy genes to determine the spatial distribution of major cell classes, and to infer their proportions in various AI domains. While excitatory neurons (SLC17A6+ cells) were distributed uniformly, there were fewer inhibitory neurons (GAD2+ cells) in RA compared to the rest of the AI (Fig. 3.1E,F). As for non-neuronal cells, which are less characterized in zebra finches, RA exhibited higher proportions of oligodendrocytes (PLP1+ cells), and to a lesser extent of astrocytes (ASS1+ cells) and microglia (RGS10+ cells) compared to the rest of the AI (Figure 3.1E,F).

*Molecular architecture of excitatory neurons in AI*

Among neuronal cells, cluster Ex_3 was defined by many genes unique to this cluster (Fig. 3.2A), several of which (e.g. GABRE, DCN) known to be differentially expressed in RA only and not in AI

or other arcopallial domains [22, 25, 38]. *In situ* hybridization for a gene with high specificity to this cluster, SRD5A2, showed a distribution restricted to RA (Fig. 3.2B), further supporting this cluster as representing an excitatory cell type unique to RA. In contrast, genes specific to other excitatory clusters (Fig. 3.2D) had low to no expression in RA but high expression in the AI outside RA (Fig. 3.2E; Supplemental Fig. 2A-C). Furthermore, several genes that are known negative markers of adult male RA (data from ZEBrA) were found to be unique to or enriched in Ex_1/2/4 (Supplemental Fig. 2C), further supporting the notion that the corresponding cell types are rare or absent in RA.

Ex_3 markers thus represent the first comprehensive characterization of genes selectively expressed or highly enriched in RA projection neurons. SRD5A2 encodes a reductase that converts testosterone into the more potent dihydrotestosterone and may play a significant role in the masculinization of RA and induction of singing behavior [7, 138, 139]. Our findings, for the first time, demonstrate that expression of this enzyme within the arcopallium is specific to RA, and assign it to excitatory neurons unique to this song nucleus. A bioinformatics analysis of the top 100 most enriched genes in this cluster found enrichment for pathways including Wnt signaling, ECM receptor interaction, and dopaminergic signaling (Supplemental Fig. 2E). Several other genes encoding proteins that are known components of the Wnt:SFRP complex (FRZB, WNT7B, WNT5B) were also unique to this cluster (Supplemental Fig. 2F), and RSPO3, also enriched in Ex_3, has been implicated in Wnt signaling regulation and associated with specification of motor neuron phenotypes [140, 141]. These findings suggest that unique Wnt signaling in RA excitatory neurons may contribute to their differentiation from other AI

excitatory neurons, or to modulation of synaptic plasticity in these cells [142]. Notably, some of the most robust Ex_3 markers were also positive markers of pre-motor song nucleus HVC (Supplemental Fig. 2H-I), a major source of input into RA suggesting a co-evolution of molecular markers within the direct vocal-motor pathway of the song control circuitry.

Access to data from excitatory clusters specific to and/or absent from RA allowed us to assess the specificity of previously identified RA excitatory cell subtypes [133]. We found that many of the markers associated with those putative RA excitatory subtypes, identified without an assessment of areas adjacent to RA, are associated with clusters in our dataset that are low in RA and enriched elsewhere in the AI (Supplemental Fig. 2D). We conclude that those clusters are not unique to RA and instead likely represent excitatory cells with broader AI distributions. This finding demonstrates the importance of anatomical context in defining gene expression specializations, and suggests that the previously identified markers of excitatory RA subtypes may be representative of broader motor subtypes rather than RA-specific cell types.

Song nuclei undergo major changes in morphology and gene expression during the developmental song learning period [42, 88, 143]. In male RA, major increases occur in volume, as well as in the size of neuronal soma and dendritic arborizations, whereas female RA undergoes atrophy, with decreases in volume and cell size, as well as cell loss. To explore whether markers that define the RA-unique (Ex_3) cluster might be developmentally regulated during this period, we next analyzed our dataset of markers of this excitatory cluster with a bulk RNA-seq dataset that profiled the developmental emergence of sex differences in RA gene

expression [44]. Many genes unique to Ex_3 and that are positive markers of RA in adult males increased during the song learning period in males only (Fig. 3.2C). This suggests that this RA-specific cell type emerges in development, concomitantly with the sexual differentiation of RA. Furthermore, many of the genes that characterize other excitatory clusters (Ex_1/2/4) and that are negative markers of adult male RA decreased in expression in male RA during the song learning period (Fig. 3.2F), suggesting a developmental regression and possible loss of these excitatory cell types within male RA.

We further discovered that the top growth factor-related genes in the developmental RNA-seq dataset that are upregulated in male RA only were highly specific to Ex_3 (Fig. 3.2G). Notably, the top gene in this set (ENSTGUG00000013568), an apparent LRRC32 paralog (which we refer to as LRRC32-2), is also the second most highly enriched gene in Ex_2 (Fig. 3.2A). Close examination of avian genomes revealed that this gene results from a songbird-specific duplication of LRRC32 (Fig. 3.2H). Compared to LRRC32-1, which shares the upstream synteny with the parent gene (Fig. 3.2H), LRRC32-2 preserves the downstream synteny, and thus more likely gained different upstream promoter regulatory elements. Only LRRC32-2 is upregulated in RA (Fig. 3.2I), suggesting it is under different epigenetic control stemming from the upstream region that diverges from the parent gene resulting in a specialization of projection neurons in RA. These findings provide an example of how two paralogs resulting from a lineage-specific duplication differentiate, with one paralog taking on novel functions. IGF2, another growth factor-related gene, is also developmentally upregulated in male RA and unique to Ex_3 (Fig. 3.2G) and its known distribution is consistent with expression selective to RA projection

neurons [144], thus possibly supporting a growth role specific to this cell type. Overall, our strategy of differentiating regional cell specializations resulted in the identification of distinct excitatory cell types with diverse distributions in the AI, as well as novel markers specific to RA projection neurons, many of which developmentally regulated and possibly critical for the differentiation and function of this cell type.

The reported caudodorsal (Cd) to ventrorostral (Vr) topography with regards to RA's projection targets [20, 145] suggests a possible molecular topography within RA. While a thorough examination of the spatial distribution in RA has been limited to a few genes, known differential markers of RA show mostly a homogenous pattern of labeling [34, 38], although limited recent evidence suggests possible molecular differences along the Cd to Vr axis [133]. To further examine that possibility, we microdissected the Cd and Vr portions of RA from thick parasagittal sections through the core of RA (Supplemental Fig. 3A, B) and performed bulk RNAseq. We did not find differentially expressed genes; in fact the two regions were remarkably similar in gene expression (Supplemental Fig. 3C). While RA cells that project to either of its major known targets (i.e. midbrain nucleus DM and medullary nXIIts/RAm) may be too sparse to have been represented in our dissections, it is also possible that these populations are not as spatially segregated as suggested by tract-tracing studies, noting that the full extent of these projections may not have been described yet. While studies on projection-specific gene expression are needed, our data suggests that RA may be more spatially homogenous in gene expression than previously suspected.

*Inhibitory neuron subtypes in the AI*

The four inhibitory neuronal clusters identified by the expression of GAD2 (In_1-4; Fig. 3.1C) differed by numerous markers selectively enriched in each of these clusters (Fig. 3.3A, expanded in Supplemental Fig. 4A). Mapping the expression patterns of cluster-specific markers by *in situ* hybridization revealed differences in the spatial distribution of these clusters within AI (Fig. 3.3B-C, Supplemental Fig. 4B). Notably, VWC2+ cells (In_1) had high densities in AId and AIv, but were very sparse in RA, likely contributing to the lower overall density of GABAergic cells in RA compared to the rest of the AI. Furthermore, In_1-4 seem to have different developmental origins based on their differential expression of markers related to embryonic ganglionic eminence subdivisions (Supplemental Fig. 4C), with In_1 and In_3 likely derived from the medial (MGE), In_2 from the caudal (CGE), and In_4 from the lateral (LGE) ganglionic eminence. In_1-4 could also be distinguished by classical markers of cortical inhibitory neurons in mammals [146] (Supplemental Fig. 5A). We observed that mammalian inhibitory markers that are more highly expressed in superficial cortical layers [101] (SNCG, LAMP5) had very low expression in In_1-4 and thus seem to be only sparsely represented in the AI, whereas markers that are more evenly distributed across cortical layers (PVALB, VIP, SST) or deep layer biased (MEIS2) were more highly expressed in the AI, with levels differing across In clusters (Supplemental Fig. 5A,B). This is consistent with the notion that the avian AI, or at least some of its neuronal populations, is analogous to deep cortical layers in mammals. Interestingly, PVALB, a classical marker of cortical inhibitory neurons in mammals, was expressed in both excitatory and inhibitory neurons in RA and AId (Supplemental Fig. 5C,D), consistent with the high density of PVALB+ cells in RA (and other song nuclei) from mRNA and protein profiling [22, 34, 38, 147,

148]. PVALB is thought to enable fast spiking in mammalian interneurons and may subserve a similar function in RA excitatory projection neurons, which were recently shown to have markedly high spiking capabilities [45].

Based on GAD2 *in situ* hybridization and consistent with previous reports [22, 149], we observed two broad GABAergic morphological types in RA, namely cells with large or small somata (Fig. 3D, right). Both cell types were also observed in other pallial song nuclei, and in all cases the large cells were also larger than GAD2+ cells outside of the song nuclei (Supplemental Fig. 6). We sought to determine if these two GABAergic cell morphotypes existed in the juvenile RA prior to the song learning period or if the difference in cell size arises during vocal development. GAD2+ cells in RA appeared uniformly small in 20 dph males (Fig. 3.3D, left). Quantitative analysis showed that adult RA GAD2+ neurons were larger than at 20 days old as well as larger than the GAD2+ cells in the caudal arcopallium (Fig. 3.3E,F). Importantly, analysis of In_1-4 markers revealed that the NPY+ cells (In_2) also included both large and small subtypes, whereas cells labeled for all other cluster-specific markers were uniformly small. The large NPY+ cells in RA were also larger than in the AI outside of RA suggesting a subset of NPY+ cells enlarge in RA during vocal development (Fig. 3.3G-I). These finding suggest that the large GABAergic cell type in RA corresponds to a subtype of NPY+ cells.

Little is known about the physiology of RA inhibitory cells, as they are difficult to access due to their sparseness. Inhibitory neurons in RA been characterized as fast spiking interneurons (FSI) with narrow action potential (AP) halfwidth and small soma [89]. Given the molecular and

morphological diversity of inhibitory neurons we observed, we sought to determine if the electrophysiological profiles of inhibitory neurons were also diverse or if they conformed to the FSI categorization. To make targeted recordings of inhibitory neurons, we injected AAV-mDlx-eGFP [150], which selectively labels forebrain inhibitory neurons, into the RA of adult males. As with *in situ* hybridization, we observed both large and small GFP-labeled cells (Fig. 3.4A). We then measured passive properties of 24 GFP+ neurons, including the membrane time constant and input resistance, via negative current injections in the whole cell current clamp configuration (Fig. 3.4B, top and bottom). We next calculated their capacitance ($C_m$) and obtained values in the ~10 to 80 pF range (Fig. 3.4C). When plotting these values in a histogram we observed two distinct peaks that could be fit with a double Gaussian (Fig. 3.4D), likely reflecting the differences in soma size observed by *in situ* hybridization and confocal imaging of GFP+ cells (Fig. 3.3D and 4A). Consistent with the imaging data, we encountered a greater number of putative small cells (small $C_m$) than large cells (large $C_m$) in our recordings.

We also measured spontaneous and evoked APs and performed a principal component analysis on passive and active membrane property measurements. Cells with high $C_m$ clustered tightly together, whereas cells with low $C_m$ were more diverse and spread out (Fig. 3.4E). This difference may reflect the molecular identification of a single large cell type compared to multiple small cell types. Putative small cells were more spontaneously active than putative large cells (Fig. 3.4F). Importantly the waveforms of the APs in all neurons, regardless of $C_m$ values, were highly reminiscent of those of previously described RA interneurons, including narrow halfwidths and sharp slopes of the rising phase following the peak of the after-

hyperpolarization (Fig. 3.4G; Spiro et al., 1999, Liao et al., 2011, Miller and Brainard, 2017). When superposing all recorded AP waveforms, there was considerable variability and no significant differences between putative large and small neurons (Supplementary Fig. 7). Additionally, the firing rates during spontaneous and evoked firing were highly variable for interneurons in both groups (Supplemental Fig. 6B-C), compared to the regular, periodic firing typical of RAPNs [45] (Supplementary Fig. S7). Interestingly all inhibitory subtypes identified were similar in expression of ion channel genes that affect membrane excitability (Supplemental Fig. 9). However, upon injecting 100 pA of current, putative small interneurons fired APs at higher frequencies than putative large ones (Fig. 3.5G-I). These findings provide evidence that GABAergic neurons in RA exhibit molecular, morphological, and electrophysiological diversity. Furthermore, the traditional classification of FSI likely corresponds to just a subset of inhibitory neurons in RA rather than being all encompassing.

*Non-neuronal cell specializations*

Non-neuronal cells also exhibited specializations in RA. For example, several other known oligodendrocyte markers in mammals were unique to the PLP1-defined cluster (Fig. 3.5A), and UGT8 and other candidate oligodendrocyte markers in ZEBrA [22] with a similar expression pattern as PLP1 were also highly specific to this cluster (Fig. 3.5A). *In situ* patterns revealed UGT8+ cells inside RA were small, with scant strongly labeled cytoplasm (Fig. 3.5B), similar to the high densities of UGT8+ cells in OM fiber bundles and tract (Fig. 3.5C, bottom right). These cells were also prevalent along fiber tracts connecting song nuclei (Fig. 3.5C, bottom left),

consistent with the heavily myelinated song system projections [151], and in sensory layer 4 thalamo-recipient analogs (Supplemental Fig. 2C), which are also heavily myelinated [152]. Besides RA, a preponderance of labeled cells for UGT8+ and other genes unique to (or highly enriched) in the PLP1-defined cluster was also seen in the other pallial song nuclei compared to adjacent brain areas (Fig. 3.5C, top). These findings strongly support this cluster as representing oligodendrocytes, and indicate that oligodendrocytes are enriched in pallial song nuclei.

Astro_1-2 clusters shared expression of classical astrocyte markers (e.g. EGF, FGFRL1), whereas Astro_2 exhibited many additional marker genes absent in Astro 1 (Fig. 3.5D). Expression of a shared Astro_1-2 marker was broadly distributed throughout the AI (Fig. 3.1E), whereas expression of CRISPLD1, one of the most highly differential genes in the Astro 2 cluster, was restricted to RA and absent in AId (Fig. 3.5D). Also notably, cells expressing CLIC4, a gene found to be highly enriched in the endothelial cell cluster, were also at a much higher proportion in RA compared to the rest of the arcopallium (Fig. 3.5E,F).

Examination of the developmental RNA-seq dataset revealed that genes upregulated in both male and female RA during the song learning period were associated with the oligodendrocyte cluster (Fig. 3.6A). These results suggest oligodendrocyte-related functions as an intrinsic feature of RA development, regardless of sex and vocal development. We confirmed this finding by performing *in situ* hybridization for PLP1, which showed a developmental increase in this oligodendrocyte marker in both male and female RA, indicating increased density of this cell type in both sexes (Fig. 3.6B). A developmental increase in myelination has previously been

described in males [151], but these results indicates that such an increase is not specific to males. Lastly, genes associated with the endothelial cell cluster increased in expression during development only in male RA (Supplemental Fig. 4C). This suggests a higher capillary density to meet a higher metabolic demand is characteristic of adult male RA compared to females and juveniles, as previously suggested [153] [44].

*Insights into RA physiology and function*

To gain insight into physiological specializations of AI cell types, we closely examined our dataset for gene families involved in fundamental neuronal processes like intrinsic excitability, synaptic transmission, connectivity, molecular transport systems, and transcriptional regulation. Several genes showed differential cell-level expression, providing first time evidence of their cell type specificity in RA and in AI more broadly (Supplemental Fig. 10). This includes various ion channel genes and neurotransmitter receptors expressed primarily in excitatory (e.g., KCNK10, CACNA2D4, CACNA1H) vs inhibitory (e.g., CACNA1I, KCND2) cells, noting the selective expression of KCNJ5 in the RA-specific excitatory cluster Ex_3, and of other channels in different excitatory and inhibitory cells (KCNH4 in Ex_1, CACNG5 in In_2, KCNS3 in In_3 – a robust positive marker of RA based on ZEBrA data, and KCNH5 in In_4) (Supplemental Figure 10A-C). We also note the predominance of various GABAergic and serotonergic receptor classes in excitatory cells, the primary expression of different dopaminergic and cholinergic receptors in excitatory (DRD3, CHRM5) vs. inhibitory (DRD2) cells, and the selective expression of GABRE in Ex_3, and of GABRG1 and CHRM2 in In_3, indicating different synaptic specializations of

various AI cell types. Intriguingly, both alpha and beta subunits of the glycinergic receptor were expressed in In_3/4, indicating that elements required for glycinergic transmission are present in these cells. As for axonal guidance and connectivity, several genes were differential in excitatory (e.g. ADCYAP1, SEMA3F, CDH17, DCN) vs. inhibitory (e.g. SLIT2, RELN, SEMA3A/C/E, UNC5C) cells (Supplemental Figure 10D). We note the selective expression of DCN in Ex_3 and RELN in In_1/3, consistent with these genes being robust positive and negative markers of RA respectively (*in situ* patterns in ZEBrA).

We also examined transcription factors (TF), focusing on those with defined DNA binding motifs and that are developmentally regulated in RA during the song learning period (Friedrich et al., 2022). Several TFs were differential across cell types (Supplemental Figure 10E-F), noting the selective expression of NKX2-8 in Ex_3, ZBTB7C and ETV4 in In_2, HEY2 and PAX6 in astrocytes, NKX6-8 in oligodendrocytes, and LEF1 in endothelial cells. Intriguingly, the speech-linked FOXP2 gene was primarily associated with inhibitory neurons and endothelial cells, not principal excitatory cells. Immediate early genes (EGR1, CFOS, CJUN) were mostly expressed in non-neuronal cells, consistent with the birds being quiet and unstimulated around sacrifice and supporting this dataset as constitutive rather than reflecting activity-inducible expression in AI. Lastly, as expected, several transporter genes (SLCs) showed expression restricted to neurons vs astrocytes (the latter for metabolites like succinate, citrate, or bicarbonate), or in excitatory (SLC17A6) vs inhibitory (SLC32A1) cells (Supplemental Figure 10G-H). We note the selective expression of a glycine transporter (SLC6A9) in Ex_2, consistent with possible local glycinergic transmission, and of the glucose transporter (SLC2A1) in endothelial cells. The various cell

types identified in this study thus exhibit unique molecular features associated with fundamental aspects of neuronal physiology, connectivity and regulatory control. These findings provide an important basis for future studies using neurophysiological, pharmacological and molecular genetic tools to assess gene expression function in RA.

*Interactive apps for data exploration*

To enable further exploration of RA cell properties, we have developed apps that allow the rapid reconstruction of cell type expression and sex-developmental data from RA single cell and bulk RNA transcriptome datasets, and integrated them with spatial expression data from the ZEBrA brain gene expression atlas (www.zebrafinchatlas.org) (Fig. 3.7C,D). This allows one to examine the brain distribution of genes selectively expressed in specific cell types (Fig. 3.7A, top) or developmentally regulated in one or both sexes (Fig. 3.7C, bottom), or conversely the RA cell type specificity or developmental profiles of marker genes in ZEBrA. As examples, genes expressed in Ex_1/3/4 (GABRA5) or in all neuronal cell types (FLRT2) show expression in RA, but those unique to clusters Ex_1/4 (CACNA1H) or In-1/3 (RELN) are restricted to the AI outside of RA (Figure 7C), consistent with Ex_3 being RA-specific but Ex_1/4 and In-1/3 being absent in RA. We also integrated the single cell and bulk RNA seq developmental transcriptome datasets by plotting how the top cluster-defining markers are regulated in the developmental sex and age contrasts (Figure 3.7B). We found that numerous adult markers of GABAergic subtypes and of non-RA excitatory neurons were downregulated developmentally in young males, their expression becoming female-biased by DPH 50, whereas several markers of the RA-specific

Ex_3 showed a developmental increase (Figure 3.7D, bottom). As for non-neuronal cells (Figure 3.7D, top), markers of the RA-specific astrocyte and of endothelial cells increased primarily in males, whereas oligodendrocyte markers were markedly upregulated in both males and females, in agreement with PLP1 (Fig. 3.6B). The incorporation of these interactive apps to the ZEBrA atlas allows the rapid examination of cell type specificity, developmental regulation, sex differences, and brain distribution. These integrated resources thus enhance the opportunities for data mining and formulation of testable hypothesis on gene regulation and function within the zebra finch vocal control system.

**Discussion:**

Understanding the function and evolution of brain circuits that control specific behaviors requires knowledge of the constituent cell types and molecular features of these circuits within their neuroanatomical context. We presented here a characterization of molecularly defined cell classes and their distribution in the AI of zebra finches, as well as identified molecular specializations linked to basic functions such as excitability, connectivity and transcriptional regulation within the identified cell classes. The AI is a complex avian brain region broadly considered analogous to deep layers of mammalian cortical areas. In songbirds it includes song nucleus RA, whose projection neurons are analogous to pyramidal cells in human laryngeal motor cortex, as well as the adjacent AId and AIv, considered analogous to deep layers of mammalian motor and auditory cortex, respectively. Several of the identified cell classes were broadly distributed in the AI, however our spatial analysis provides strong support for the existence of specialized cell types across AI subdivisions. This includes an excitatory neuron, a

large inhibitory subtype, and an astrocyte type unique to RA, as well as evidence of an intriguing high enrichment of oligodendrocytes in RA (Figure 3.8). These data advance our understanding of the physiology and evolution of RA within the song control circuitry, as well as of AI more broadly.

Including both RA and other parts of the AI in the dissections provided neuroanatomical context to our cellular transcriptome analysis. This approach, previously used in an extensive regional analysis of RA markers by *in situ* hybridization [38], allowed the identification of cellular and molecular specializations truly unique to RA. While we identified fewer inhibitory clusters than in Colquitt et al. [133], which had the different goal of defining subtypes of cells within RA, we were able to identify previously undescribed regional molecular differences in excitatory cells and astrocytes, and regional specializations of broadly distributed GABAergic subtypes. Because Colquitt et al. [133] did not assess the arcopallium outside RA, the context of their identified cell types was not addressed, and it is thus unclear which and/or how many are truly unique to RA (Supplemental Fig. 7D).

Furthermore, integrating our single cell data with the bulk RNA-seq datasets from Friedrich et al. [44] revealed that many of the cell type markers of adult RA are developmentally regulated during the song learning period in a sex-specific manner. In particular, many of the markers of the RA-specific Ex_2 are upregulated developmentally in males only, whereas markers of cell types absent in adult male RA are downregulated in males and/or upregulated in females during this same period. These findings strongly suggest that excitatory cells developmentally

differentiate themselves during the song learning period, in a sex-specific manner, with important implications for understanding the origins of RA as a specialized song control nucleus. Our data provide clear evidence that the major projection neuron in RA (Ex_2) is a cell type that expresses unique molecular markers not expressed in neurons in other AI regions. The identification of these markers addresses an important knowledge gap, as the majority of known RA markers are negative ones [22]. Genes selectively expressed in RA projection neurons likely reflect molecular functions that are active within these cells and are thus poised to affect their growth, differentiation and functions. This includes mediation of spectral features of song, through patterned firing [18]. Zebra finch RA is highly dimorphic, its large projection neurons subserving song production in males, whereas it is regressed in females, which do not sing [7]. The growth of RA and other song nuclei, as well as singing behavior, is largely mediated through sex steroids [138]. The enzyme encoded by SRD5A2, one of the most differential markers of Ex_3, converts testosterone into 5-alpha-dihydrotestosterone (DHT), which potentiates the actions of other sex steroids and is required for the full masculinization of RA. SRD5A2 is one of the most specific RA markers to date by in situ, and it is developmentally upregulated in males only. Its cell specificity indicates that RA projection neurons are the site of local DHT production, which can then act directly on androgen receptors. Other genes related to growth and differentiation were also selectively expressed in Ex_3, several of which developmentally upregulated in male RA only, and could potentially contribute to RA's growth and differentiation. This includes a novel paralog of a growth factor gene (LRRC32) that is a key regulator of the TGF-β pathway, a TF linked to early development of the nervous system (NKX2-8)[154], and several genes linked to Wnt signaling.

During development, the axons of RA projection neurons are likely guided by a combination of attractive and/or repulsive guidance molecules to establish their unique connections [20, 74]. Identifying the cell specificity of axonal guidance genes provides well-informed hypotheses on the regulation of projection neuron connectivity in RA and other AI regions. Furthermore, while parts of the zebra finch pallium and mammalian cortex may be divergent in their origin [133], their convergent expression of markers can give clues to function of specific populations of arcopallial neurons. For instance, KLHL14, highly enriched in Ex_3 cells, was recently identified as expressed in a specific population of mouse primary motor cortex neurons, its deletion causing axons to extend into the spinal cord past their brainstem targets. KLHL14 could thus be an important determinant of specific connections of Ex_3 to the brainstem. ADCYAP1, in contrast, enriched in AI excitatory clusters that predominate outside of RA and a robust negative RA marker is a marker of deep layer corticopontine/corticospinal neurons in the mouse [155] consistent with the projections and function of the AI, and in particular AId [83]. Intriguingly, SLIT2 was exclusively expressed in inhibitory neurons in the AI, whereas SLIT1, ROBO1, and ROBO2 were found in non-RA excitatory cells, but predominantly in RA inhibitory cells. While some aspects of SLIT-ROBO axon guidance signaling, especially the lack of SLIT1 in RA, might be involved in long-range connectivity, these findings seem inconsistent with an involvement of SLIT-ROBO signaling in establishing the connections between RA and nXIIts/RAm [99]. Instead, our data suggest that SLIT-ROBO mechanisms are more likely involved in establishing local inhibitory networks.

There is evidence that GABAergic transmission is important for regulating RA spiking behavior [89, 102], but our knowledge of inhibitory cells in RA has been limited. Our study considerably expands this knowledge, including substantial evidence of features unique to RA, even though the proportions and molecular identities of inhibitory neuron types may be largely conserved across brain regions and subregions [133]. For instance, we found the lower density of GABAergic neurons in RA compared to the rest of the AI to be largely due to the lack of In_1. Furthermore, inhibitory neurons in RA have been previously described as small, fast-spiking interneurons. While we did observe small cells, we found sparser, larger GABAergic cells, that do not conform to the small fast-spiking interneuron classification. These large GABAergic cells originate as a subset of NPY+ cells that undergo marked growth during the vocal learning period. They seem to be also characteristic of other pallial vocal nuclei both in songbirds and in other vocal learners, and may be a fundamental feature shared by vocal learning circuits. We show conclusive evidence that a large proportion of the large PVALB+ cells in RA are in fact excitatory, another unique feature of pallial vocal nuclei, this one possibly linked to their very high metabolic demands.

We also show novel and extensive data on the expression profiles and spatial distribution of major non-neuronal cell classes within the zebra finch arcopallium, noting that non-neuronal cells in the avian brain have been largely understudied. We observed multiple non-neuronal specializations in RA that could impact vocal production. One example was an astrocyte subtype unique to RA. The two astrocyte subtypes were remarkably similar with the apparent expression of a subset on genes in the RA subtype. Astrocytes have many known roles that

impact synaptic transmission suggesting these astrocytic specializations may impact local transmission within RA [156]. We also observed a preponderance of oligodendrocytes within RA as well as the other pallial song nuclei compared to their surrounds. Like in the mammalian brain, oligodendrocytes are most commonly found in fiber tracts of the zebra finch brain [22]. Oligodendrocyte enrichment within the pallial song nuclei suggests that some local axonal projections intrinsic to these nuclei are myelinated. This unusual pattern, possibly originating from collaterals of longer-range axonal fibers, may reflect a specialization to support high neuronal firing rates within these nuclei [18]. Also noteworthy is the male-specific developmental increase in markers of endothelial cells, suggesting a possible expansion of the capillary beds to meet the high metabolic demands of RA.

RA has previously been hypothesized to have evolved in songbirds as a vocal-motor specialization of a more primordial AId already present in a non-vocal learning avian ancestor [35]. We recently supported this hypothesis by showing that RA and AId share molecular specializations compared to the rest of the arcopallium [38]. This observation is consistent with RA having differentiated from AId during evolution to support the production of learned vocalizations. A duplication and specialization model was recently proposed in the evolution of cerebellar nuclei [157], whereby a cerebellar nucleus is duplicated and its inhibitory neurons are maintained, while the excitatory neurons diverge in their gene expression and projection targets. By analogy, and based on the hypothesis that RA evolved as a specialization of AId [35], RA's excitatory cells have distinct connections from the rest of the AI, and were heavily divergent from other excitatory cell types in terms of molecular markers. RA's excitatory

neurons may thus have evolved through differentiation of an ancestral AI excitatory neuron, adapting expression profiles to enable vocal production. Conversely, inhibitory neurons are molecularly conserved within the AI but with size and density differences in RA, with In_2-4 more uniformly distributed throughout the AI, In_1 more sparsely distributed in RA, and a subset of In_2 cells in RA only exhibiting large soma. With respect to non-neuronal cells, RA possesses a unique subtype of astrocytes which shares a large portion of its transcriptome with non-RA astrocytes, with few key differences. Overall, this study identifies which cell classes likely diverged in the evolution of the song system and the degree to which (e.g. transcriptome, cell size, density) that divergence occurred.

**Methods**

Animals:

All procedures involving live animals were approved by OHSU's IACUC and are in accordance with NIH guidelines. Adult male zebra finches (*Taeniopygia guttata*) were isolated in a sound dampening chamber overnight and sacrificed by decapitation the next morning prior to lights on to minimize auditory, vocal, and movement-related activity dependent changes in gene expression. Juvenile zebra finch brains were prepared as previously described [38, 45].

Single nuclei isolation and sequencing:

Frontal slices (400µm thick, one slice per hemisphere) containing RA, AId, and AIv were cut on a Leica VT1200S vibratome in an ice-cold cutting solution containing 119mM NaCl, 2.5mM KCl, 8mM MgSO4, 16.2mM NaHCO3, 10mM HEPES, 1mM NaH2PO4, 0.5mM CaCl2, 11mM D-Glucose, 35mM Sucrose with a pH of 7.3-7.4 when bubbled with carbogen (95% O2, 5% CO2; osmolarity ~330-340 mOsm). Slices were then transferred to an incubation chamber containing artificial cerebral spinal fluid with 119mM NaCl, 2.5mM KCl, 1.3mM MgSO4, 26.2mM NaHCO3, 1mM NaH2PO4, 1.5mM CaCl2, 11mM D-Glucose, 35mM Sucrose with a pH of 7.3-7.4 when bubbled with carbogen (95% O2, 5% CO2; osmolarity ~330-340 mOsm) at 37°C. Microbiopsies containing RA, AId, and AIv were dissected using a scalpel and angled forceps and flash frozen in a dry ice/isopropyl alcohol slurry. Tissue was homogenized in 300uL of nuclei lysis buffer (10mM Tris pH 8.0, 250mM sucrose, 25mM KCl, 5mM MgCl2, 0.1% Triton-X 100, 0.5% RNasin, and 0.1mM DTT). 700uL of lysis buffer was added to the homogenate and the sample was incubated on ice for 5 minutes. The sample was transferred to a 15mL conical tube and

centrifuged at 500 x g for 5 minutes at 4°C. The supernatant was discarded and 1mL of ice-cold lysis buffer was added and incubated on ice for 5 minutes. The sample then was centrifuged at 500 x g for 5 minutes at 4°C. The pellet was resuspended in 1mL of nuclei suspension buffer (1x PBS, 0.01% BSA, and 0.1% RNasin), and filtered through a FLOWMI 40 µm tip strainer (Bel-Art) and centrifuged at 500 x g for 5 minutes at 4°C. The pelleted nuclei were resuspended in nuclei suspension buffer. Libraries were prepared using the Chromium Single Cell 3' Library & Gel Bead Kit v3 (10x Genomics) according to the manufacturer's instructions. Libraries were sequenced using an Illumina NovaSeq 6000 at the Massively Parallel Sequencing Shared Resource at OHSU. Reads were aligned to the zebra finch genome (Taegut1) using CellRanger with include-introns. SnRNA-seq data was processed using a standard Seurat workflow [158]. Clusters were resolved using FindClusters() with a resolution of 0.25 and visualized with a UMAP projection. Differential gene expression was defined using FindMarkers() with min.pct of 0.25. One cluster exhibited differential markers with low percent of cells (<50%) exhibiting the defining marker and with a majority of markers having a negative average logFC, suggesting low quality nuclei and was excluded from future analysis. ConsensusPathDB was used for pathway enrichment analysis. All pathways were selected with a minimum overlap of 2 and p-value cutoff of 0.05. Protein complex-based gene sets were also selected with the same criteria. We manually annotated the ensembl models that appeared in differential expressed gene sets for protein coding genes using Ensembl.org.

Bulk RNA-sequencing

The dorsal and ventral portions of RA (each an estimated 1/3 area of RA in that slice) were microdissected from a 400µm thick sagittal vibratome section (n= 4 animals), prepared as described above. The dorsal/ventral axis was defined as in alignment with the descending occipito-mesencephalic tract (OM), based on the pattern of retrogradely labeled cells in RA from previous studies [20, 74, 145]. RNA from these samples were isolated using a Qiagen RNeasy Micro Kit and cDNA libraries were generated using Takara Bio SmartSeq v4 PLUS kit. Paired end sequencing (2 x 100 bp) was performed on an Illumina NovaSeq 6000 with a target depth of 50M reads per sample.

Trimming was performed with trimmomatic (v0.36) and quality control with FastQC (v0.11.9) with no issues detected. STAR (v2.6.1) was used to align sequencing reads and generate read counts per gene using bTaeGut1_v1.p assembly (GCF_003957565.1) and associated genome features from the NCBI annotation release 104 for zebra finch.

We also analyzed an existing bulk RNA-seq dataset (Friedrich et al.). This dataset is comprised of microdissections of RA from 20dph and 50dph male and female zebra finches. Count tables were downloaded and transcript abundances were plotted using custom code (Friedrich et al.). As described in Friedrich et al, a binomial generalized linear model was fit to the data with sex and age as main factors with a sex + age interaction. Genes with a Benjamini-Hochberg-based false discovery rate (FDR) < 0.01 were considered significant.

*In situ* hybridization

*In situ* hybridizations were carried out as previously described (Carleton et al., 2014; Nevue et al., 2020). All probes had been previously described with the exception of PLP1, RGS10, NPY, COLEC12, VWC2, CRISPLD1, CERKL, and ADCYAP1. We cloned the 3' region of one gene (SRD5A2), for which there was no cDNA clone available, using PCR (F: GAGAGGTGGGAGGGTCTCAT, R: TCCATGTGTGCAGTGTGGTC), followed by a second round of PCR with a primer containing the T3 polymerase promoter. Briefly, plasmids containing cDNA of the gene of interest were isolated and restriction enzyme digested with BSSHII to release the insert. The insert was purified using a QIAquick PCR purification kit (Qiagen). Antisense DIG-tagged probes were synthesized by *in vitro* transcription using T3 polymerase (Promega) and purified using a Sephadex G-50 column. Probes were hybridized to sections overnight at 65°C followed by a series of high stringency washes. Sections were blocked an incubated in anti-DIG-AP (1:600; Roche) for two hours and incubated in BCIP/NBT chromogen (PerkinElmer) overnight. For fluorescent *in situ* hybridizations, following the high stringency washes, sections blocked and incubated in anti-DIG-POD (1:600; Roche) for two hours. Slices were then washed and incubated in Alexa 350 or 488-conjugated tyramide (1:100; Invitrogen) for two hours.

Analysis of in situ hybridization images

Images of in situ hybridizations were captured on a Nikon Eclipse E600 microscope. Proportion of cell types in each arcopallial subdivision was determined by quantifying an 200x200um square window in each subdivision. Because AIv contains a higher cell density than RA and AId [38], the values for AIv were normalized by a factor of 0.8. Area of cells were quantified using

ImageJ. Images were converted into 8-bit, thresholded to remove background and subjected to two binary processes (close- and open) to generate particles. For the images from ZEBrA [22], optical density measurements were taken in ImageJ in RA and HVC as well as the shelf of HVC (ventral to HVC) and the cup of RA (located rostro-ventral to RA).

Biophysical characterization of GABAergic neurons

To virally target GABAergic interneurons in the arcopallium, we used AAV9-mDlx-GFP (Addgene #83900; Dimidschstein et al., 2016). Stereotaxic injections of 500nL were made bilaterally in adult male zebra finches anesthetized with isoflurane. Coordinates used for RA were AP: -0.5mm, ML: 2.2-2.7mm, DV: 2.5mm. Birds were allowed to recover for 2-3 weeks before being sacrificed. Vibratome slices (200um) were prepared in the same manner as for microdissections described above. Slices were then transferred to an incubation chamber containing artificial cerebral spinal fluid (aCSF) with (in mM): 119 NaCl, 2.5 KCl, 1.3 $MgSO_4$, 26.2 $NaHCO_3$, 1 $NaH_2PO_4$, 1.5 $CaCl_2$, 11 D-Glucose, 35 Sucrose, pH 7.3–7.4 when bubbled with carbogen (95% $O_2$, 5% $CO_2$; osmolarity ~330–340 mOsm) for 10 min at 37 °C, followed by a room temperature incubation for ~30 min prior to start of electrophysiology experiments. RA was visualized using differential interference microscopy and transduced cells in RA were visualized using a 488nm filter. Whole-cell current-clamp recordings were made using a HEKA EPC-10/2 amplifier controlled by Patchmaster software (HEKA, Ludwigshafen/Rhein, Germany). Data were acquired at 100 kHz and low-pass filtered at 2.9 kHz. Patch pipettes were pulled from standard borosilicate capillary glass (WPI, Sarasota, FL, USA) with a P97 puller (Sutter Instruments, Novato, CA). All recording pipettes had a 3.0 to 6.0 MΩ open-tip resistance in the bath solution. Electrophysiology data were analyzed off-line using custom written routines in IGOR Pro (WaveMetrics, Lake Oswego, OR, USA).

Intracellular solutions contained (in mM): 142.5 K-Gluconate, 21.9 KCl, 5.5 $Na_2$-phosphocreatine, 10.9 HEPES, 5.5 EGTA, 4.2 Mg-ATP and 0.545 GTP, pH adjusted to 7.3 with KOH, ~330-340 mOsm. To initiate current clamp recordings, we first established a giga-ohm seal in the voltage clamp configuration, set the pipette capacitance compensation (C-fast), and then set the voltage command to -70 mV. We then applied negative pressure to break into the cell. Once stable, we switched to the current clamp configuration. Whole-cell current-clamp recordings were performed at room temperature as described in [45]. We note that recordings were not corrected for a calculated liquid junction potential of +9 mV. Capacitance was calculated from the measured membrane time constant and input resistance ($\tau_{m(ms)} = R_{in(M\Omega)} \bullet C_{m(pF)}$). Principal component analysis was performed using scikit-learn in a Jupyter Python environment. For the principal component analysis, any missing values were imputed by using the average value of the measurement based on the size classification.

In separate birds, following the 2-week recovery time we perfused the bird with 0.9% saline followed by 3% paraformaldehyde. Brains were dissected and post-fixed in 3% PFA overnight at 4C before being transferred to PBS. 200µm thick slices were cut on a vibratome and transferred to a well plate containing CUBIC reagent overnight [159, 160]. Slices were then mounted on slides, coverslipped, and imaged on a ZEISS LSM 980 confocal microscope.

Interactive application development

Apps to interact with the single cell and bulk RNA seq datasets used in this study were developed using the shiny R package. The apps are hosted and deployed using shinyapps.io.

Apps can be accessed on the homepage of our spatial gene expression atlas ZEBrA at www.zebrafinchatlas.org. Code to generate the shiny apps are available at www.github.com/samifriedrich.

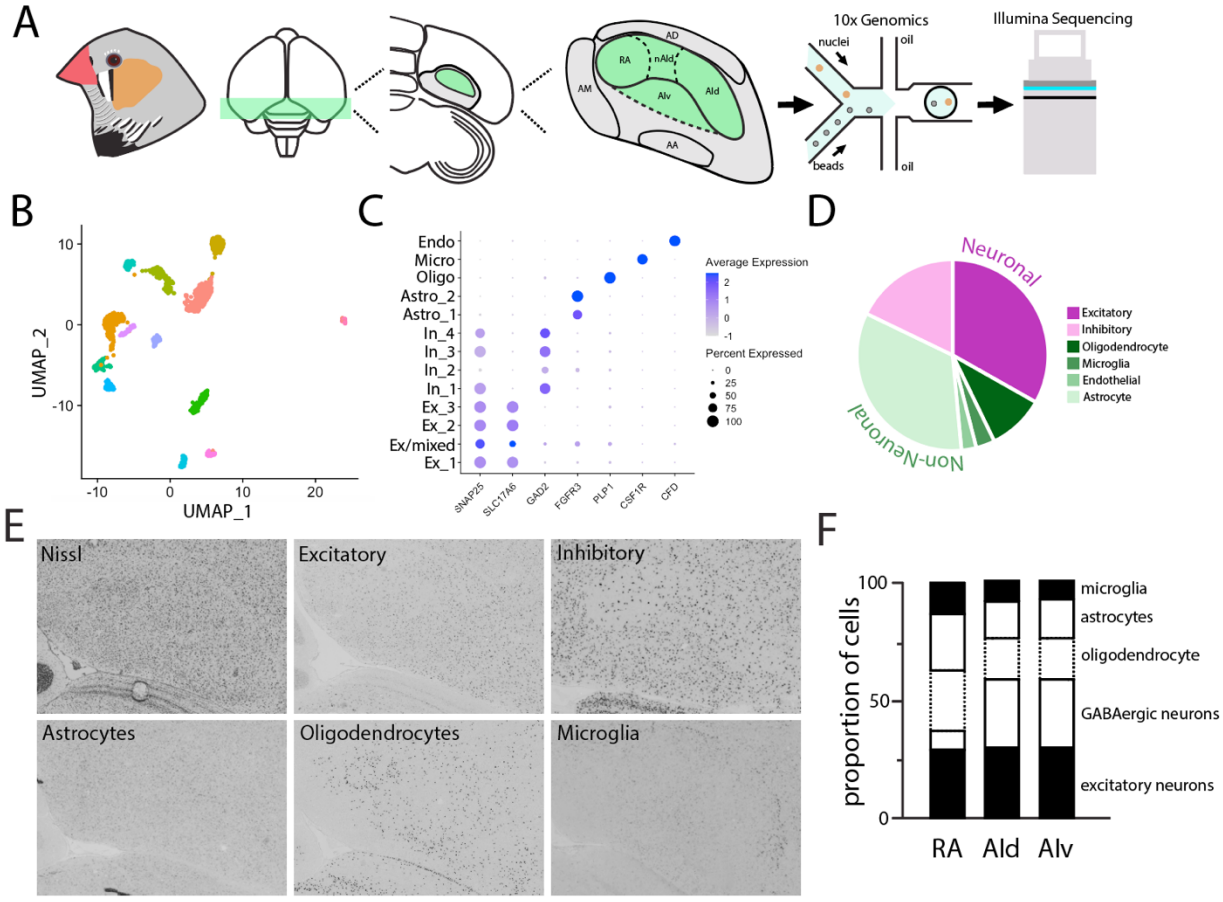



Figure 3.1: Cell classes of the intermediate arcopallium. A: Workflow to generate single nuclei transcriptomes of cells in the intermediate arcopallium. RA, AId, and AIv subdivisions were microdissected from thick vibratome sections, followed by nuclei isolation, droplet-based profiling on the 10x Genomics platform, and Illumina sequencing. B. UMAP plot visualization of cell classes in the intermediate arcopallium (N=2 animals; nuclei=1504). C. Dotplot for cluster-specific genes used for cluster identity. SNAP25, neuronal; SLC17A6, excitatory; GAD2, inhibitory; SOX9, astrocyte; PLP1, oligodendrocyte; CSF1R, microglia; FLT1, endothelial. D. Proportion of cell types from snRNAseq data present in the intermediate arcopallium: astrocytes (33.2%), oligodendrocytes (9.62%), microglia (3.19%), endothelial cells (2.52%), excitatory neurons (33.63%), inhibitory neurons (17.85%). E. In situ hybridizations for cell type proxy genes. Excitatory (SLC17A6), Inhibitory (GAD2), Astrocytes (ASS1), Oligodendrocytes (PLP1), Microglia (RGS10). F. Proportion of each cell class in the three subdivisions of the AI. Abbreviations AA, anterior arcopallium; AD, dorsal arcopallium; AM, medial arcopallium; AId, dorsal intermediate arcopallium; nAId, neck of the dorsal medial arcopallium; RA, robust nucleus of the arcopallium; AIv, ventral intermediate arcopallium.

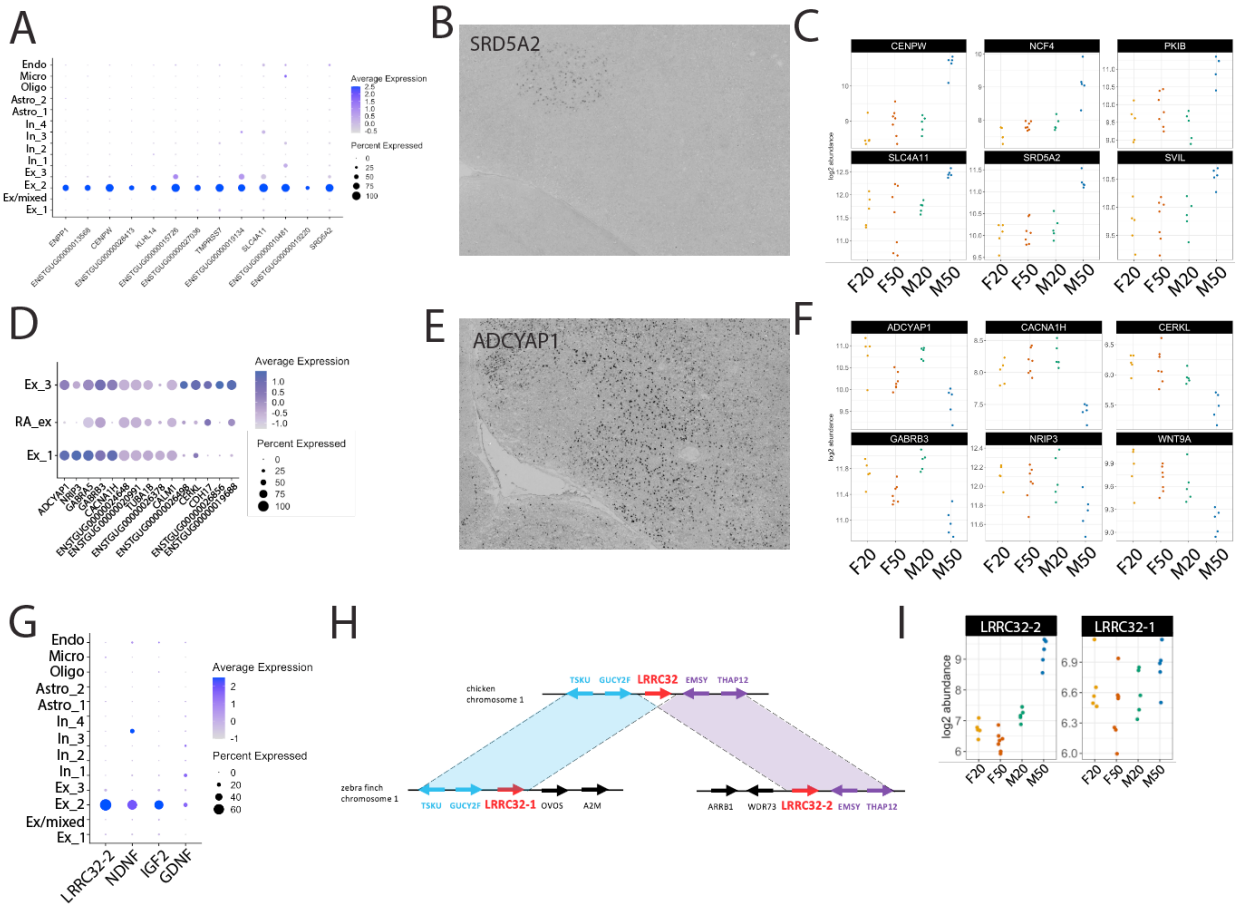



Figure 3.2: Molecular specializations of RA excitatory neurons. A. Dot plot of 25 markers of the RA excitatory neuron cluster. B. In situ hybridization for SRD5A2, one of the most differential RA markers, illustrating high specificity of this cluster. C. Bulk RNAseq plots of six RA excitatory markers that increase during song development. D. Dot plot of the top 8 genes in each of the non-RA excitatory clusters. E. In situ hybridization for ADCYAP1, one of the most differential non-RA excitatory genes. F. Bulk RNAseq transcript abundance plots of six excitatory markers absent in RA that decrease during song development. G. Dot plot for enriched growth factor genes in RA excitatory neurons. H. Genomic map of chicken chromosome 1 and zebra finch chromosome 1 exhibiting the duplication of LRRC32. I. Bulk RNAseq plots showing only one copy is developmentally regulated. All RNAseq plots are presented as FDR<0.01 for the Male 20-50 comparison and FDR>0.01 for the Female 20-50 comparison with the exception of LRRC32-1 which is FDR>0.01 for both comparisons.

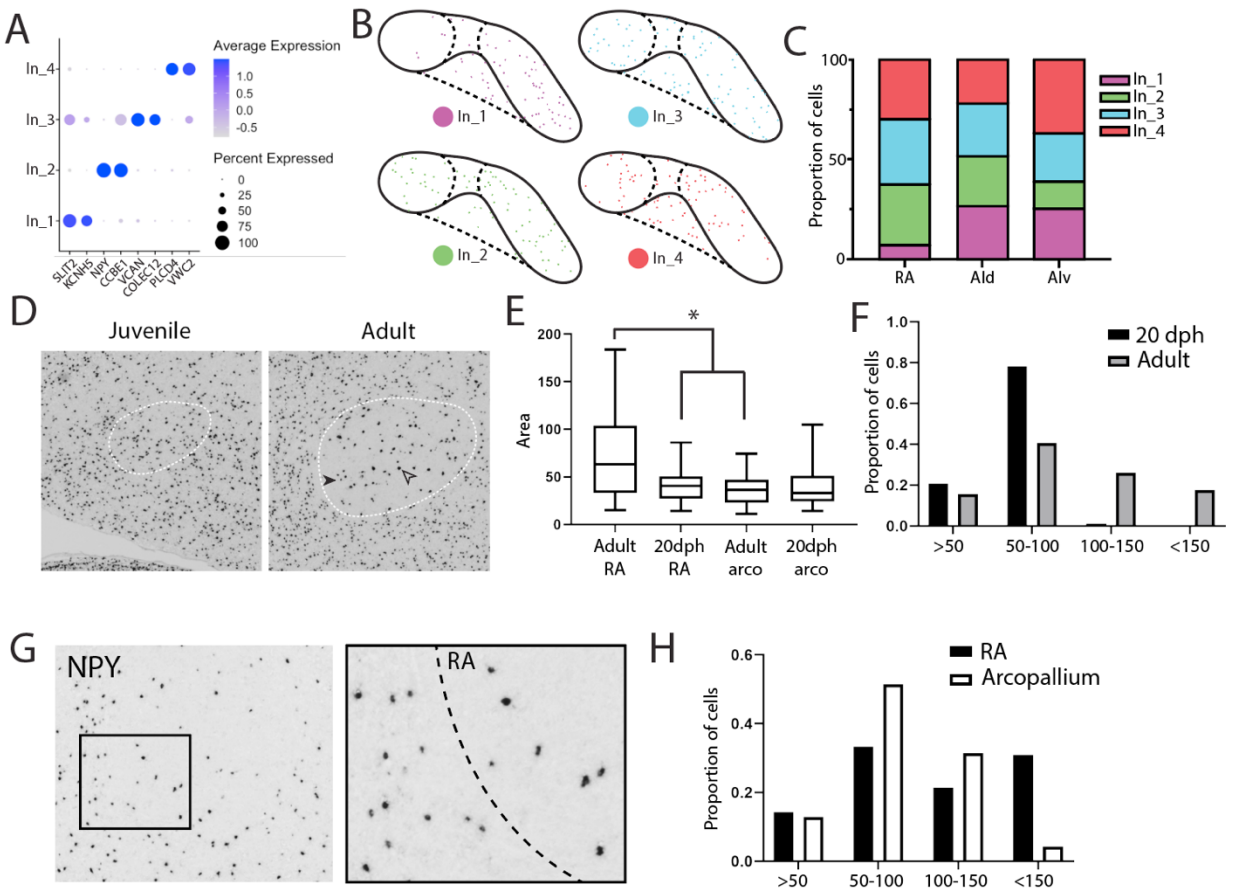



Figure 3.3: Inhibitory subtypes of the intermediate arcopallium. A. Gene markers used to differentiate the four classes of GABAergic neurons. B. Spatial distribution of inhibitory subtypes as aggregation of in situ hybridizations. C. Quantification of spatial distribution of inhibitory subtypes. D. GAD2 in situ hybridization at 20dph and Adult RA. Closed arrow denoting examples of large GAD2+ morphotype, open arrow denoting examples of small GAD2+ morphotype. E. Quantification of GAD2+ cell area in juvenile (20dph) and adult RA and caudal arcopallium. Asterisk represents <0.0001 p-value following one-way ANOVA multiple comparisons. Adult RA vs 20dph RA $p<0.0001$, Adult RA vs Adult outside $p<0.0001$, 20 RA vs 20 outside $p=0.9596$. F. Quantification of GAD2+ cell area in juvenile (20dph) and adult RA and caudal arcopallium. Asterisk represents <0.0001 p-value following one-way ANOVA multiple comparisons. Adult RA vs 20dph RA $p<0.0001$, Adult RA vs Adult outside $p<0.0001$, 20 RA vs 20 outside $p=0.9919$. G. In situ hybridization for NPY in arcopallium. Inset highlights difference in cell size within and outside of RA. H. Area distribution of NPY+ inhibitory neurons in RA and the adjacent arcopallium.

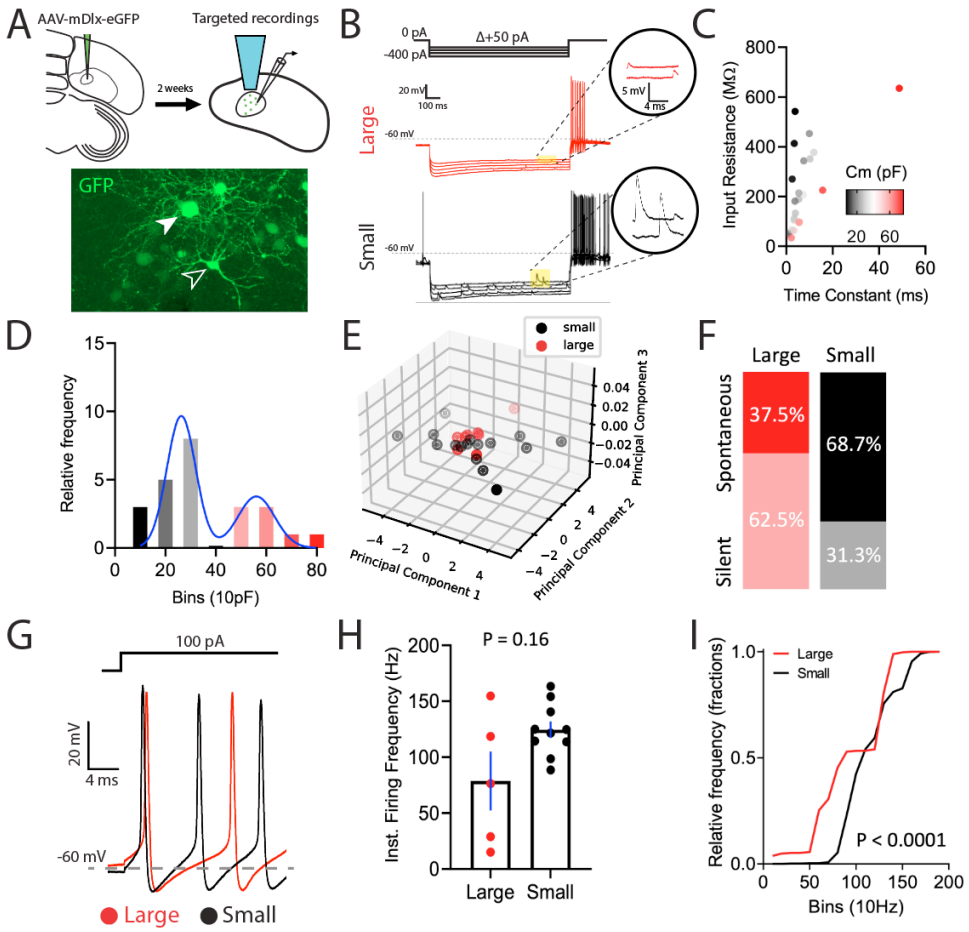



Figure 3.4: Electrophysiological identification of putatively large and small GABAergic interneurons using GFP expression under an mDlx promotor. A. Schematic of experiment. AAV-mDlx-eGFP was injected into RA. Targeted recordings of GFP+ were made. Open arrow denotes small soma, closed arrow denotes large soma neuron. B. Five overlaid example traces of negative current injections in GFP positive neurons used to calculate capacitance ($C_m$) for small (black traces) and large (red traces) GABAergic interneurons. Inset shows the difference in the size of mini EPSPs detected during current injections. Note the small cells with larger input resistances display larger potential changes. C. Colorimetric graph plotting the input resistance versus the time constant of GFP positive cells recorded. Color scale shows the $C_m$ of the individual neurons recorded. N = 24 neurons. D. Histogram of the relative frequency of calculate $C_m$ for all GFP positive cells recorded. N = 24 neurons; Bins = 10 pF. Histogram was fit by two Gaussian curves revealing distinct populations of neurons as determined by the calculated $C_m$. E. Principal component analysis for recorded cells. PC1 accounted for 31.7% of the variance, PC2 accounted for 23.0%, and PC3 13.1%. F. Proportion of cells that were spontaneously active. G. Overlay of the first 24 ms of a +100 pA current injection in neurons with large (red trace) and small (black trace) $C_m$ measurements. H. Dot plot comparing the instantaneous firing frequencies of large (red) and small (black) $C_m$ neurons. Mann-Whitney U test, U = 13. I. Cumulative frequency distributions of firing frequencies from large (red) and small (black). Kolmogorov-Smirnov test, D = 0.4126, N = 289 large cell and 1044 small cell events.

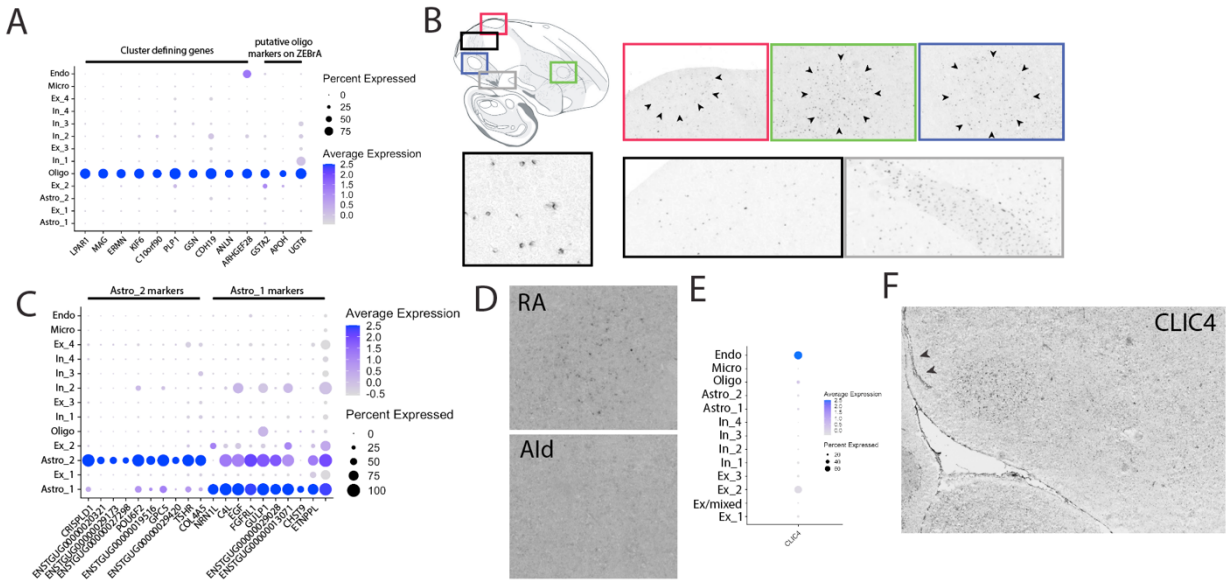



Figure 3.5: Non-neuronal specializations in song nuclei. A. High magnification of oligodendrocytes in RA showing their small size and strong labeling. B. Oligodendrocyte enrichment in pallial song nuclei HVC (red), RA (blue), and LMAN (green). Shown also is the fiber tract from HVC to RA (black) and the OM tract (grey) analogous to the pyramidal tract. Gene shown is UGT8 from ZEBrA. C. Top 10 markers for each astrocyte cluster D. In situ hybridization for CRISPLD1, an Astro_2 enriched gene. E. Dot plot for CLIC4, a gene enriched in endothelial cells. F. In situ hybridization for CLIC4. Arrows denote expression in ventricle. Note, some labeling in RA may be due to Ex_2 expression.

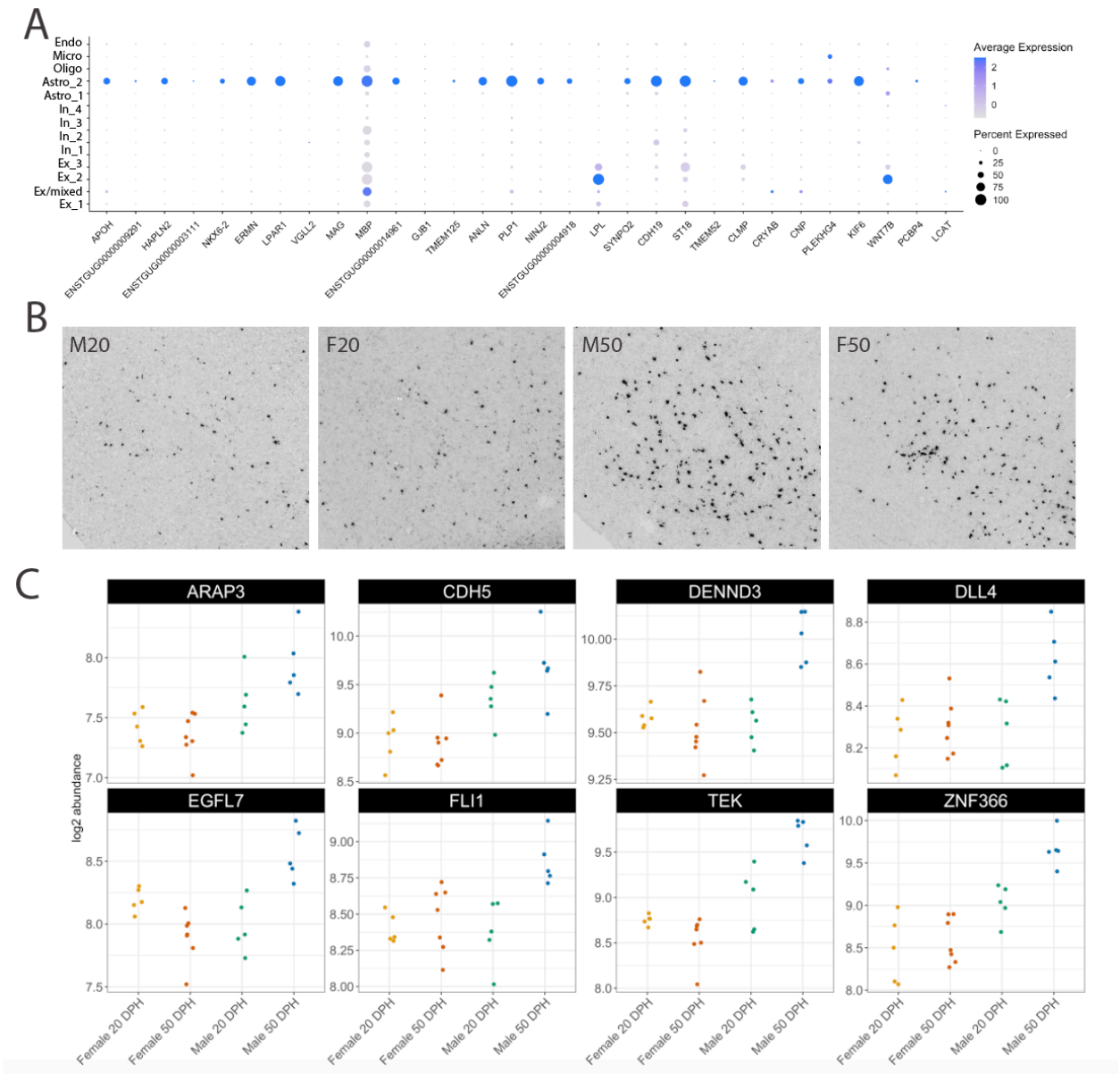



Figure 3.6: Developmental changes in non-neuronal composition in RA. A. Most enriched DEGs that increase in RA in both male and females. B. PLP1 expression in RA during development. C. Endothelial cell markers how a male-specific developmental increase in expression. All plots are FDR<0.01 for the Male 20-50 comparison and FDR>0.01 for the Female 20-50 comparison.

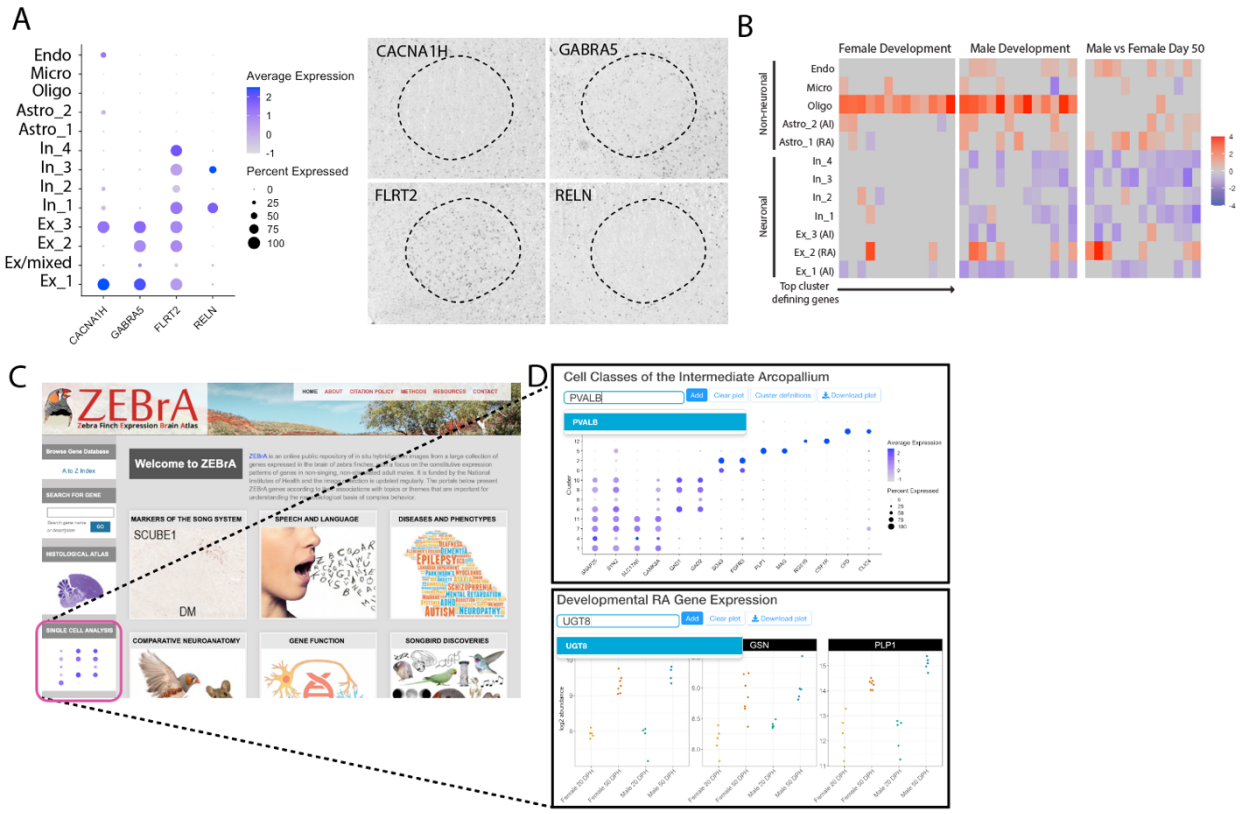



Figure 3.7: Interactive integration of spatial and transcriptomics data. A. New insights from gene expression specializations in ZEBrA using cell type data. B. Integration of cell type and developmental datasets. C. Shiny apps to interact with the datasets presented in this study are linked on the home page of ZEBrA (zebrafinchatlas.org). D. Example of usage of shiny apps.

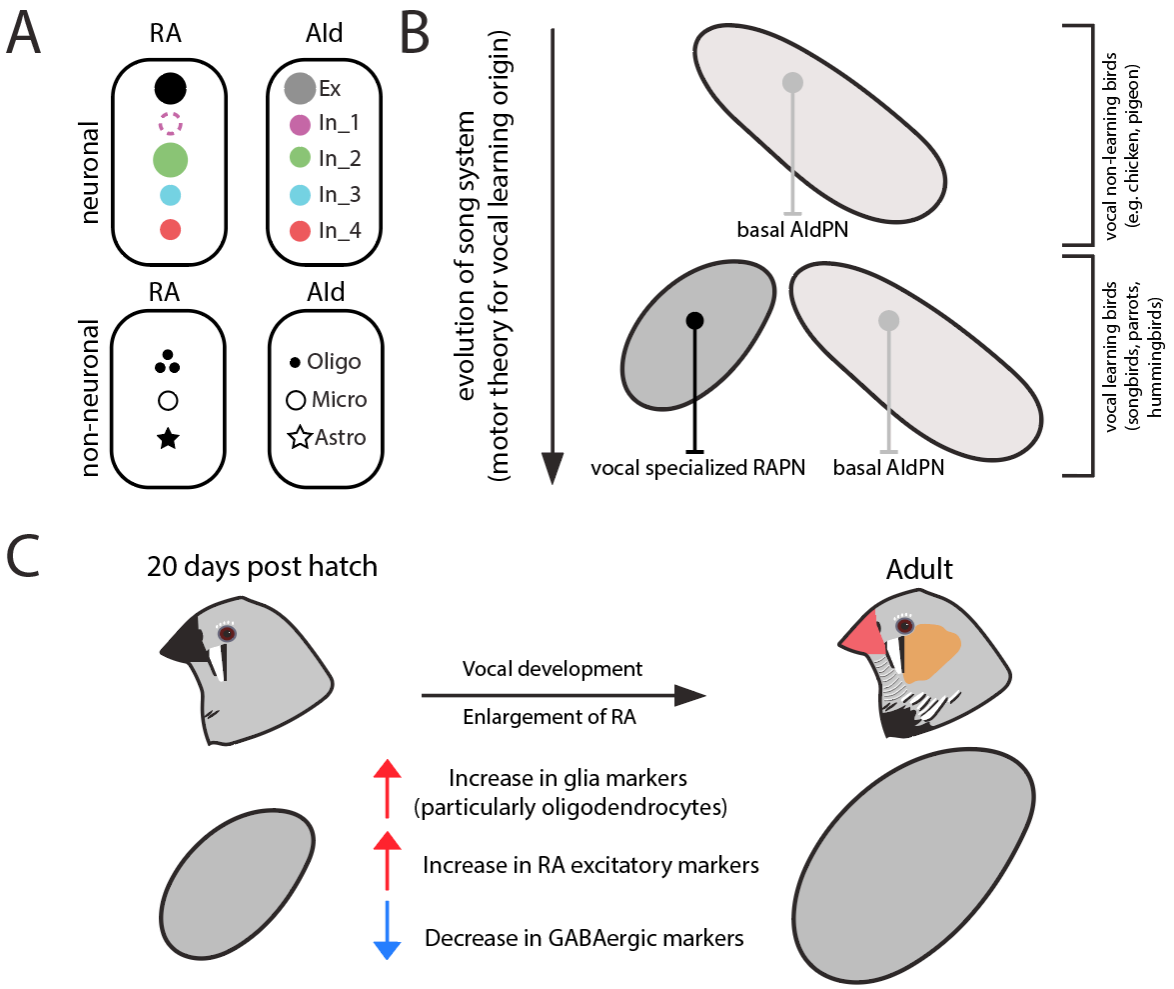
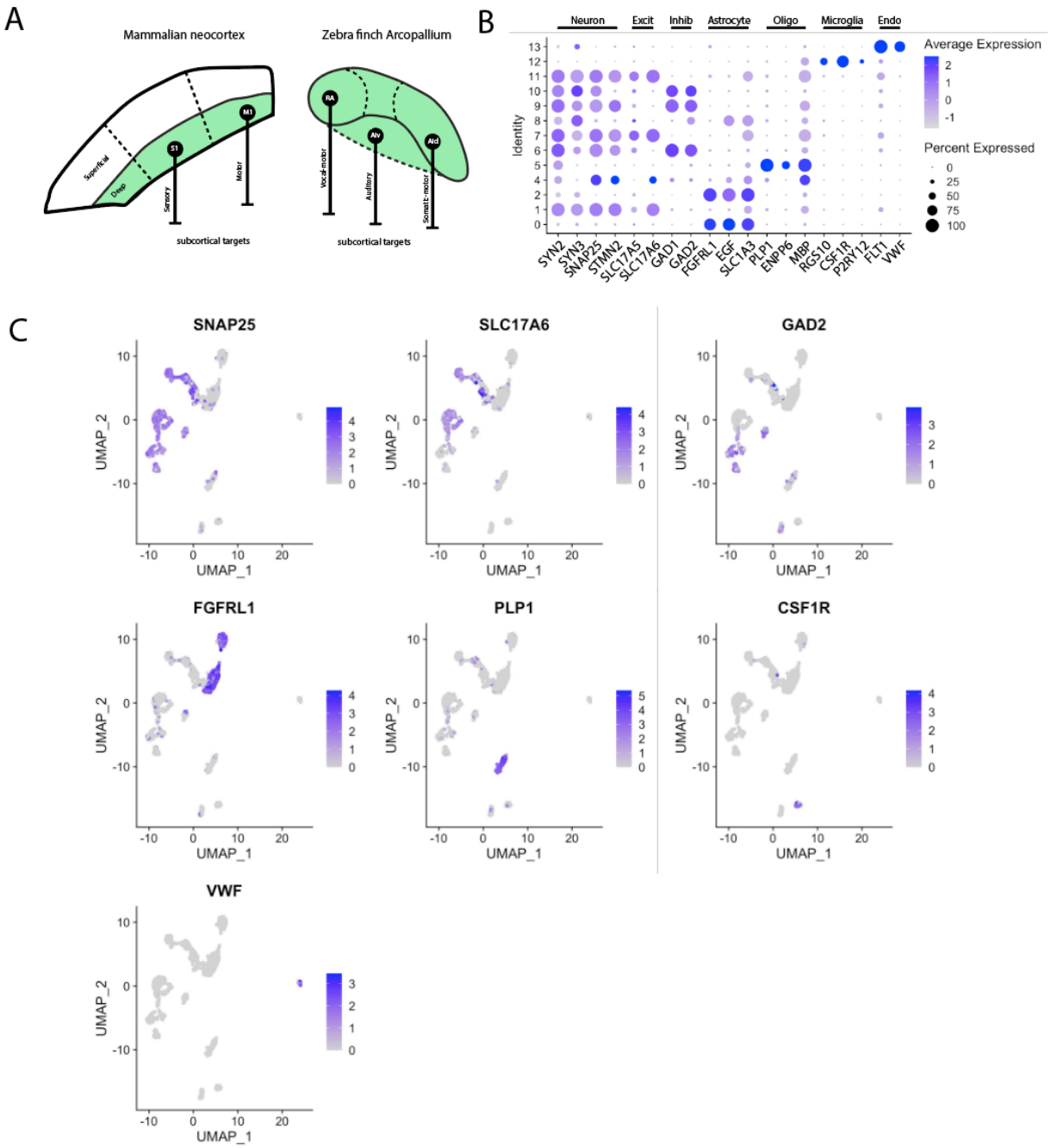
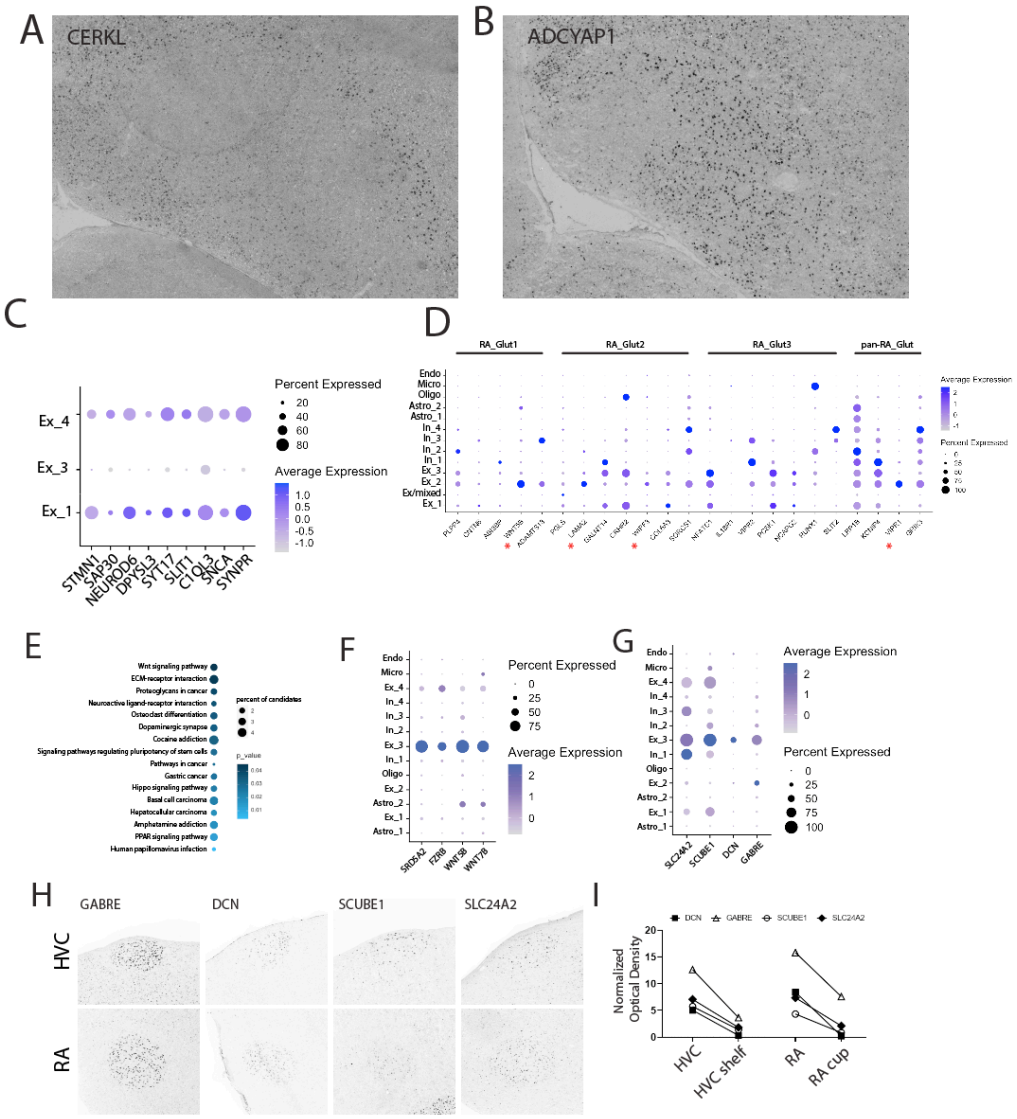


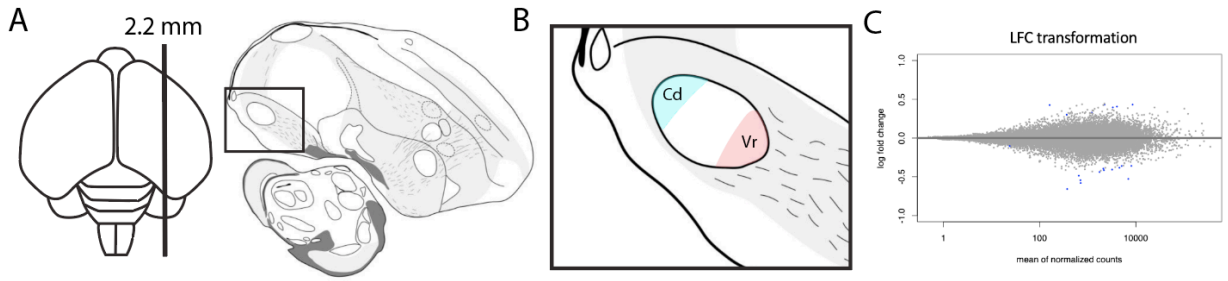

Figure 3.8: Summary of RA cell type specializations. A. RA excitatory neurons exhibited different gene expression from AId. GABAergic subtypes of AId were present in RA but with In_1 having a lower density and a subset of In_2 being large. Non-neuronal differences included oligodendrocytes were more numerous in RA and RA contained a differentiated astrocyte subtype. B. RA is hypothesized to have evolved as a specialization of the pre-existing AId. During evolution for vocal-motor specializations, RA specialized excitatory neurons (RAPN) from the existing AIdPN. C. Changes in RA cell type composition during vocal development. With an enlargement of RA comes an increase in non-neuronal markers, increase in RA excitatory markers, and a decrease in GABAergic markers. These specializations likely contribute to the physiology of producing learned vocalizations. Note, schematics are not to scale.



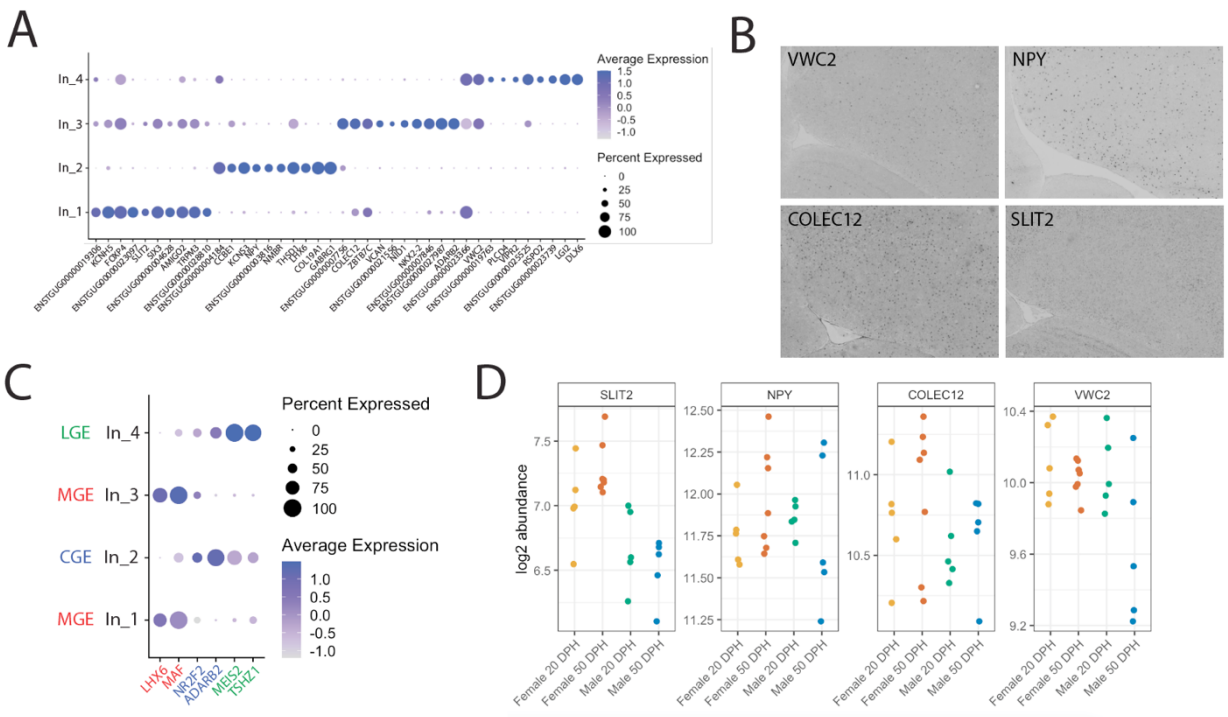

Supplemental Figure 1: Overview of intermediate arcopallium A. Schematic of comparison between descending motor and sensory projections in the mammalian neocortex and the zebra finch arcopallium. B. Dotplot for common cell type markers and differential markers used for cell type annotation and in situ hybridization spatial investigation. C. UMAP plots for each of the cell types.



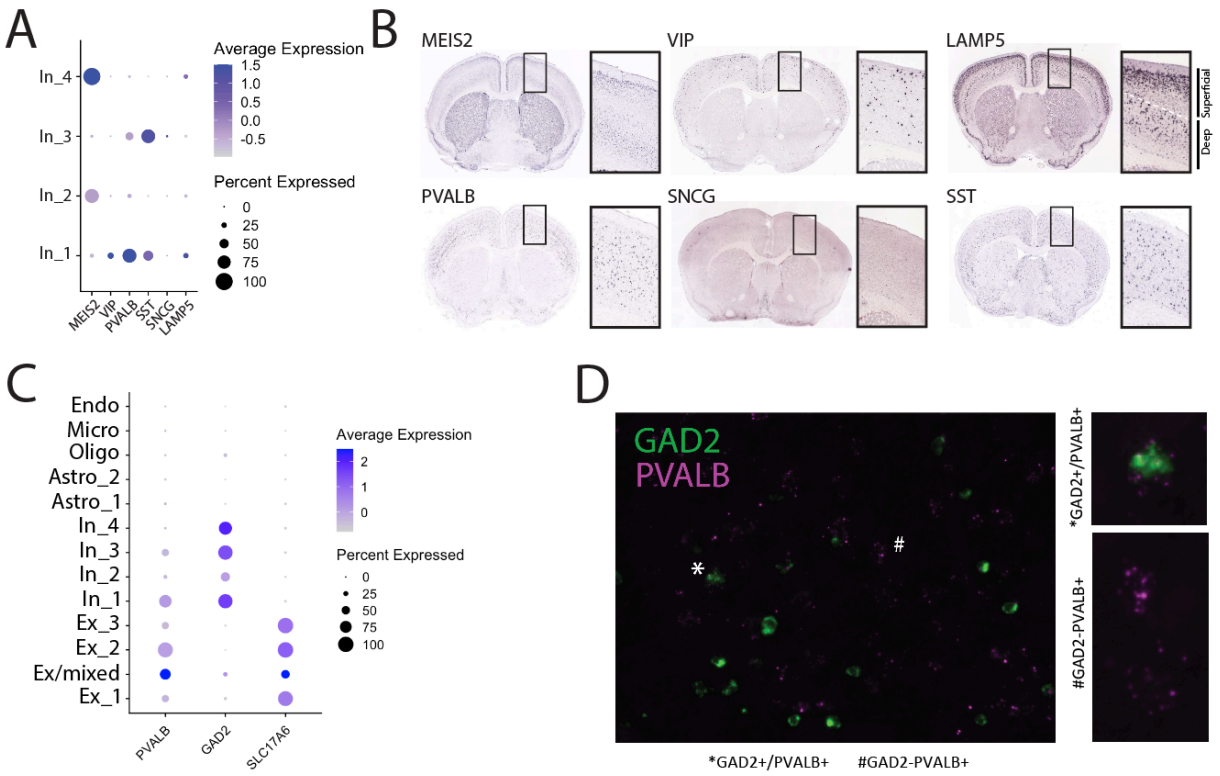

Supplemental Figure 2: Excitatory neurons of the intermediate arcopallium. A. In situ hybridizations for Ex_1 enriched gene (ADCYAP1) suggesting the cluster is pan-AI in expression B. Ex_3 enriched gene (CERKL) suggesting this cluster is AIv-biased. C. Negative RA markers from ZEBrA showing their specificity for AI excitatory clusters. D. Colquitt et al. performed sc/nRNAseq on RA (without a comparison region) and established three molecularly distinct excitatory clusters. Mapping DEGs from those clusters onto data generated in this study reveals a majority of genes are not specific to RA (e.g. present in non-RA excitatory neurons) or not specific to RA excitatory cells (e.g. present in inhibitory neurons). Red asterisks denote qualitatively unique expression in RA excitatory neurons. E. Gene Ontology for RA markers highlighting potential specialized pathways in RA. F. Dot plot of RA-specific gene (SRD5A2) and genes found in the Wnt:SFRP complex from the pathway enrichment analysis. G. Dot plot for four genes with similar expression in RA and HVC. H. In situ hybridizations from ZEBrA (zebrafinchatlas.org, Lovell et al., 2020). I. Optical density measurements for the four RA/HVC markers showing enrichment compared to surrounding region. Optical density measurements are normalized to background expression.



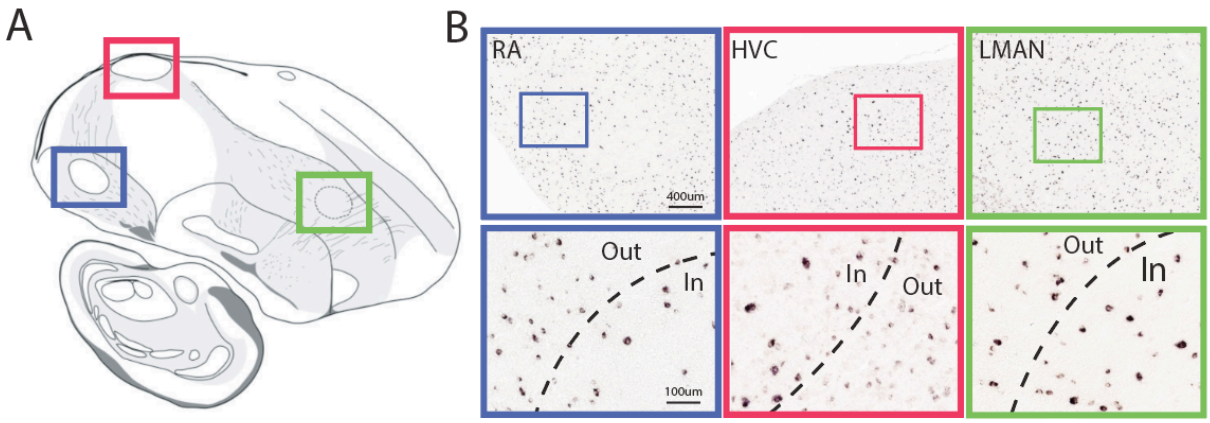

Supplemental Figure 3: Homogeneity of RA. A. Schematic representing the location of the sagittal slice used for microdissections. B. Location of the microdissections representing the caudodorsal (Cd) and ventrorostral (Vr) divisions of RA used for bulk RNAseq. C. Differential gene MA plot from bulk RNA-seq of Cd and Vr.



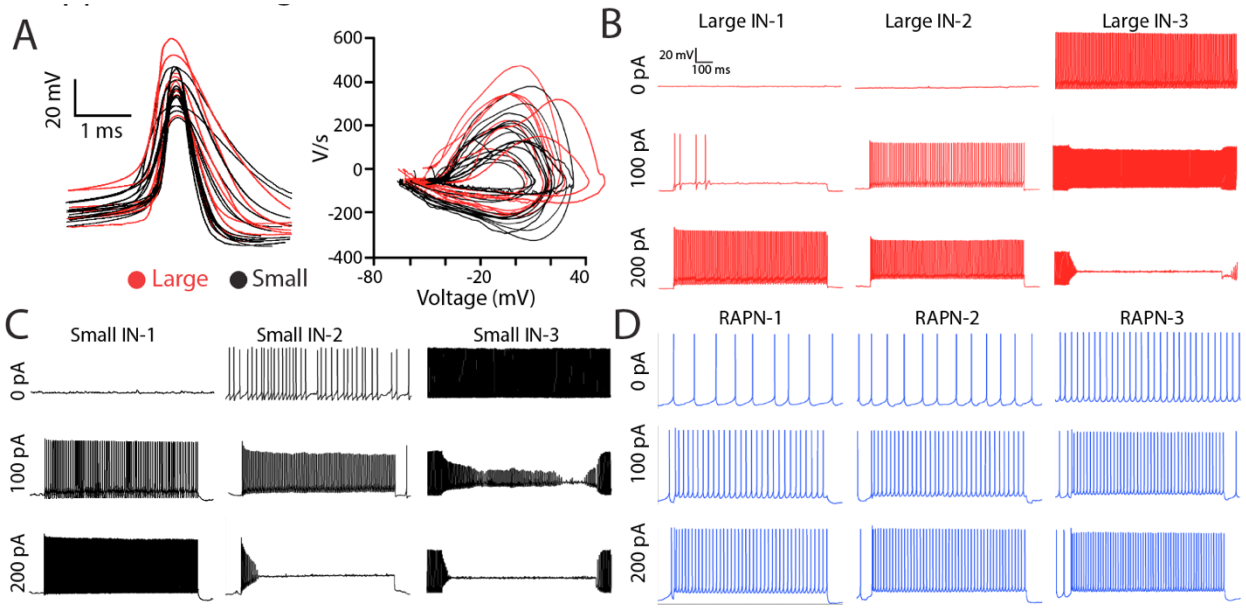

Supplemental Figure 4: GABAergic cell types in the intermediate arcopallium. A. In situ hybridization for cell type proxy genes representing the four inhibitory subtypes in the intermediate arcopallium. B. Dotplot of top 10 markers for each of the four inhibitory subtypes. C. Dotplot of ganglionic eminence markers to assign developmental origin to inhibitory subtypes. D. Bulk RNAseq transcript abundance plots for cell type proxy genes for each of the four inhibitory subtypes.



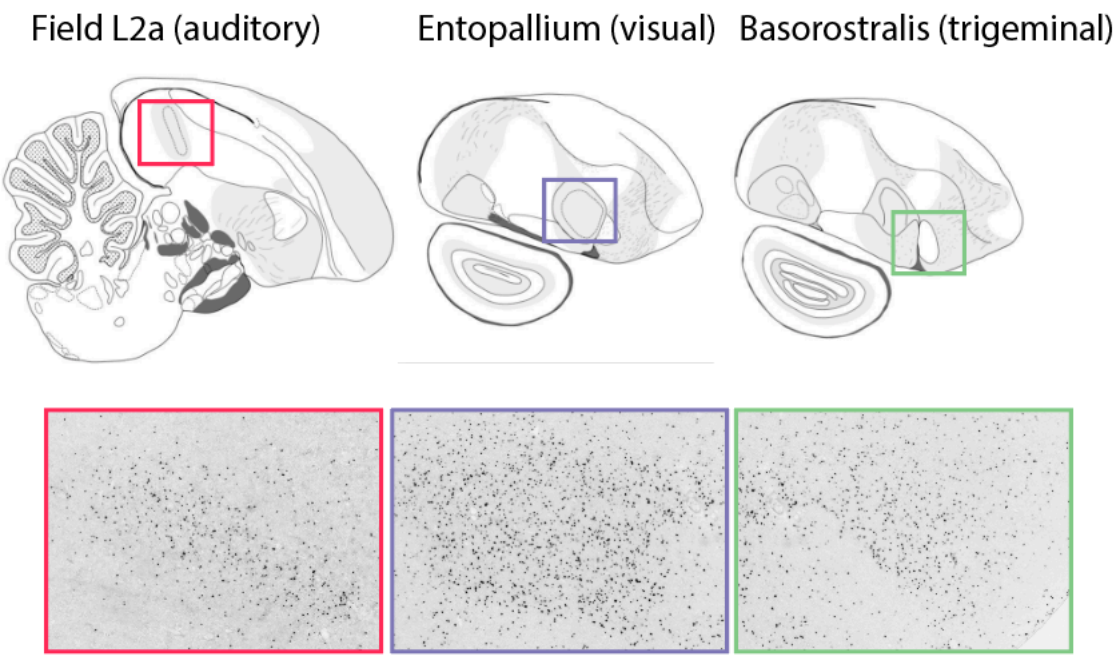

Supplemental Figure 5: Unique gene expression in inhibitory subtypes. A. Dotplot for classical inhibitory markers (Bakken et al., 2021 Nature). B. Screenshots of mouse in situ hybridization from the Allen Brain Atlas (Lein et. al). C. Dotplot for genes of interest with expression in inhibitory neurons. D. Dual fluorescent in situ hybridization of PVALB and GAD2 showing co-labeled cells, and PVALB+/GAD2- cells.



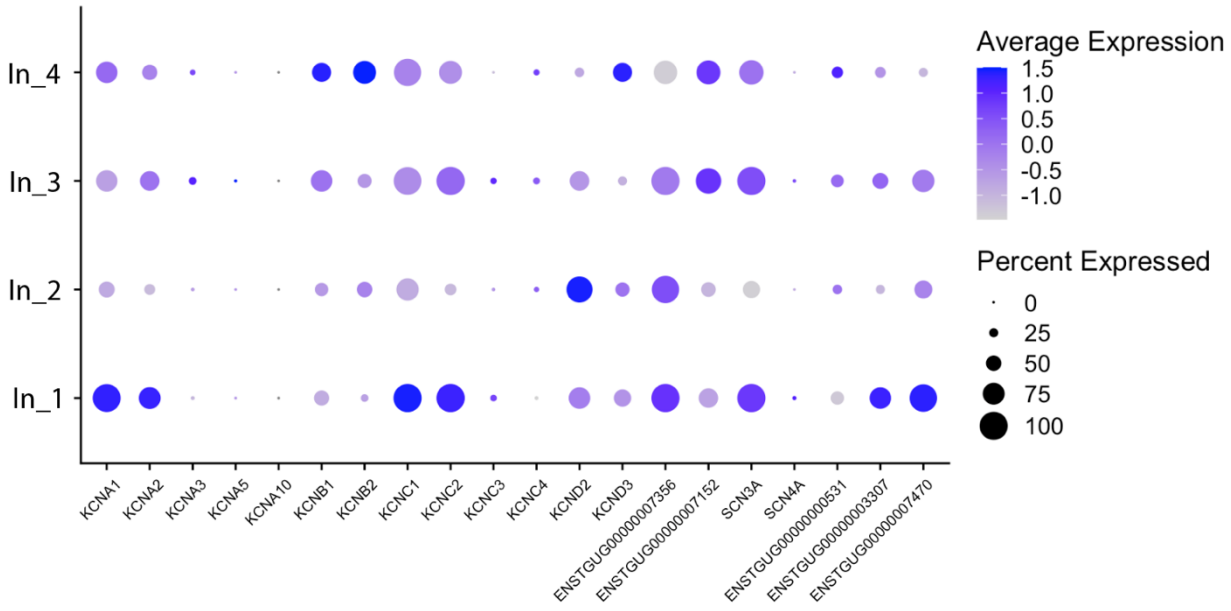

Supplemental Figure 6: Large GABAergic cells are an evolutionary feature of the song system. A. Sagittal schematic of the song system highlighting the three pallial song nuclei. B. GAD2 in situ hybridization images in RA, HVC, and LMAN. Insets (below) show sparser, larger GAD2+ cells in RA and HVC compared to the adjacent arcopallium and nidopallium, respectively. GAD2+ LMAN cells are of similar density compared to the surrounding nidopallium with a subset being larger.



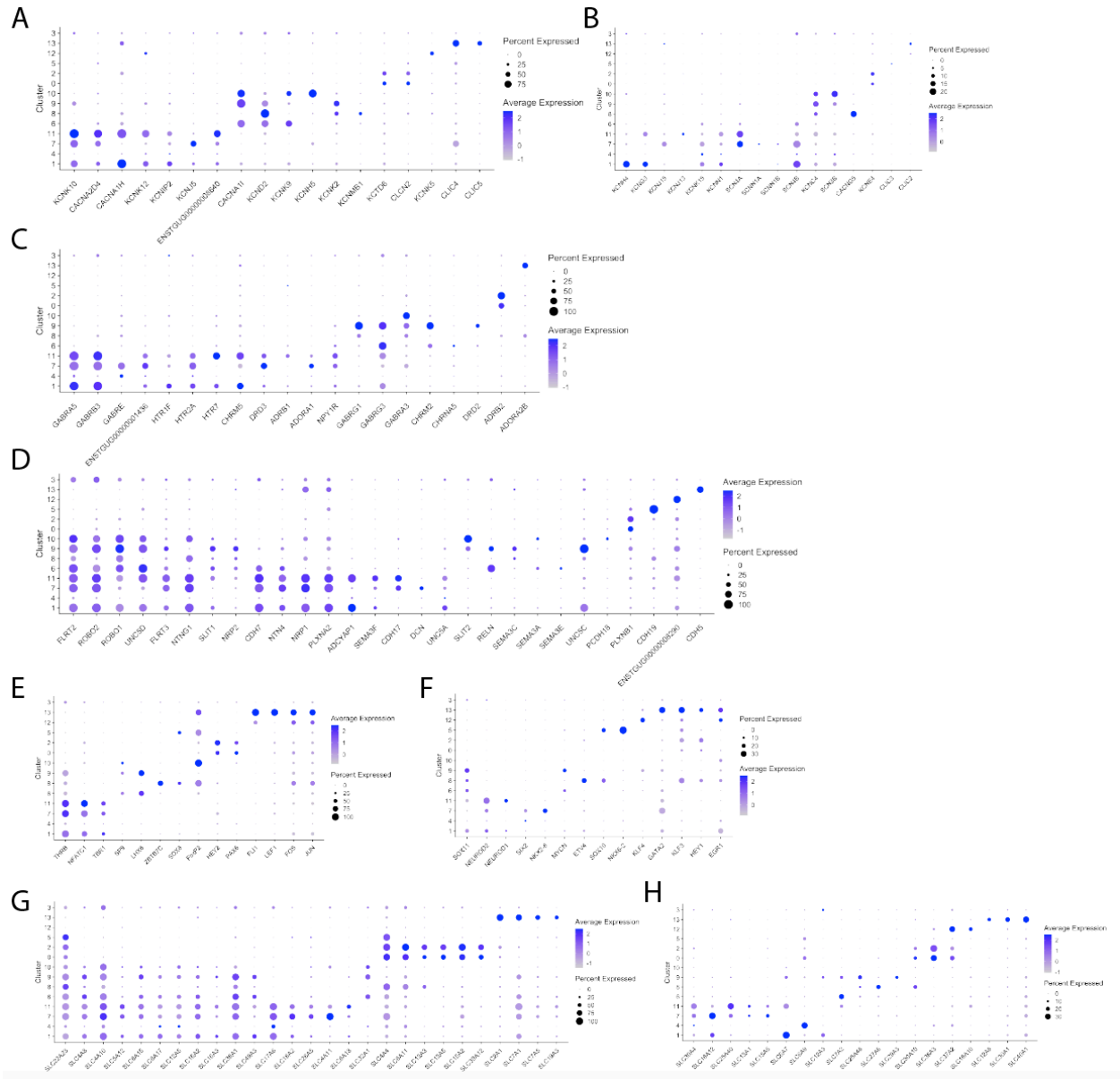

Supplemental Figure 7: Examples of mDlx+ neuron recordings. A. Overlays of action potential waveforms (left) and phase plot (right) for putatively large and small neurons. B. Three putatively large neurons with different spiking behavior at 0 pA, 100 pA, and 200 pA current injection. C. Three putatively small neurons with different spiking behavior at 0 pA, 100 pA, and 200 pA current injection. D. Three examples of RA projection neuron spiking behavior at 0 pA, 100 pA, and 200 pA current injection.




Supplemental Figure 8. Oligodendrocyte enrichment in layer 4 cortical analogs.




Supplemental Figure 9: Expression of intrinsic excitability genes in inhibitory neuron classes. Dot plot for voltage gated potassium channel family members Kv1-4 and voltage gated sodium channel alpha subunits. ENSTGUG00000007356 is SCN1A; ENSTGUG00000007152 is SCN2A; ENSTGUG00000000531 is SCN5A; ENSTGUG00000003307 is SCN8A; ENSTGUG00000007470 is SCN9A.




Supplemental Figure 10: Expression of physiologically relevant gene families. A. Ion channel genes with enrichment and high expression in clusters. B. Ion channels with enrichment in clusters but with few cells highly expressing the gene. C. Neurotransmitter receptors with enrichment in clusters. D. Axon guidance related genes with enrichment in clusters. E. transcription factors with enrichment and high expression in clusters. F. Transcription factors with enrichment in clusters but with few cells highly expressing the gene. G. Solute carrier genes with enrichment and high expression in clusters. H. Solute carrier genes with enrichment in clusters but with few cells highly expressing the gene.

# Chapter 4

## Sodium and potassium ion channel specializations in the songbird intermediate arcopallium

Portions of this chapter were published as *Zemel BM, Nevue AA, Dagostin A, Lovell PV, Mello CV, von Gersdorff H. Resurgent Na+ currents promote ultrafast spiking in projection neurons that drive fine motor control. Nature Communications, 2021. 12(1), 1-23* and *Zemel BM, Nevue AA, Tavares LE, Dagostin A, Lovell PV, Jin DZ, Mello CM, von Gersdorff H. Cortical Betz cells analogue in songbirds utilizes Kv3.1 to generate ultranarrow spikes. biorxiv*. AAN collected and analyzed the *in situ* hybridization and genomics data, BMZ collected and analyzed the electrophysiology data. AAN, BMZ, CVM, HvG wrote the text.

**Abstract:**

The underlying mechanisms that promote precise spiking in upper motor neurons controlling fine motor skills are not well understood. Here we report that projection neurons in the adult zebra finch song nucleus RA display robust high-frequency firing, ultra-narrow spike waveforms mediated by the voltage gated potassium ion channel Kv3.1, superfast Na+ current inactivation kinetics, and large resurgent Na+ currents. These properties of songbird pallial motor neurons closely resemble those of specialized large pyramidal neurons in mammalian primary motor cortex. They emerge during the early phases of song development in males, but not females, coinciding with a complete switch of Na+ channel subunit expression from Navβ3 to Navβ4. We

show that the adult RA exhibits unique expression of KCNC1 which enables the ultra-narrow spikes compared to the adjacent arcopallium. Together, these ion channel expression specializations help facilitate the complex behavior of song production.

**Introduction:**

The speed and accuracy of muscle control and coordination depend on the spiking activity of upper motor neurons [161]. In mammals, a subclass of layer 5 pyramidal neurons (L5PNs) project from the motor cortex to various targets in the brainstem and spinal cord. These cells are involved in specific aspects of fine motor control and they produce narrow half-width action potentials (APs) [162, 163]. Changes to their intrinsic properties have been implicated in facilitating the learning of complex motor skills in some species [121, 164-167]. Notably, primates and cats possess varying numbers of very large L5PNs with wide-caliber myelinated axons, fast AP conduction velocities, and ultra-narrow AP spikes [121, 168, 169]. These specialized Betz-type cells, first discovered in the human motor cortex, send projections that often terminate directly onto lower motor neurons[166], and are thought to be involved in highly refined aspects of motor control [121, 161, 168, 169].

Birds display a diverse array of complex behaviors and cognitive skills ranging from elaborate nest building and tool usage to episodic memory and vocal mimicry [56, 170, 171]. Remarkably, birds accomplish this without a typical six-layered neocortex, which underpins the capacity for complex motor skills in mammals [13, 37]. Nonetheless, avian pallial nuclei form microcircuits that appear analogous to those in the mammalian neocortex [13, 14, 24, 37, 172]. In songbirds, the robust nucleus of the arcopallium (RA) plays a key role in singing and provides direct

descending projections to brainstem motor neurons that innervate the avian vocal organ [11, 43, 57] and respiratory muscles [173]. RA projection neurons (RAPNs) can thus be considered analogous to L5PNs in the motor cortex [14, 24, 37, 172]. Indeed, these cells share important features, like wide-caliber myelinated axons and multiple spine-studded basal dendrites [89], although RAPNs lack the large, multilayer-spanning apical tufted dendrites that are a hallmark of L5PNs [174-176]. However, detailed knowledge of the ion channel composition, biophysical properties, and firing patterns of RAPNs is limited. Therefore, it is still unclear to what extent they function in an analogous manner to L5PNs.

Zebra finch RAPNs face considerable spiking demands during singing, which requires superfast, temporally precise coordination of syringeal and respiratory musculature [177-179]. As the adult male sings, RAPNs exhibit remarkably precise spike timing (variance ~0.23 ms) [180]. RAPNs also exhibit increased burstiness during the developmental song learning period, their instantaneous firing rates changing from 100 to 200 Hz when they produce immature vocalizations (subsong) to 300–600 Hz when a song becomes mature (crystallized). Average overall spike rates of RAPNs increase from 36 Hz at subsong to 71 Hz in adults [91]. Song maturation thus correlates with reduced variability in the timing of increasingly high-frequency bursts with a refinement of single spike firing precision. Importantly, nerve firing rates of >75 Hz are required for force summation in the superfast syrinx muscles [181]. This high spike frequency in RAPNs during song production is energetically demanding as indicated by increased staining for cytochrome C in maturing males, but not female zebra finches [153].

Songbird RAPNs thus seem to share with Betz-type L5PNs similar evolutionary pressure for fast and precise signaling, a constraint that can lead to neurons in unrelated species sharing similar expression patterns for a specific repertoire of ion channels. A good example is electric fish species from different continents, which have convergently evolved electrocytes exhibiting similar physiology and molecular features [182, 183]. Indeed, high-frequency firing electrocytes with extremely narrow AP spikes co-express fast activating voltage-gated $Na^+$(Nav) and $K^+$ (Kv) channels [184, 185]. We hypothesized that RA projection neurons were molecularly specialized compared to other arcopallial neurons to facilitate high-frequency firing. Here we explored the molecular specializations of Nav and Kv channels in RA. We correlate molecular specializations with biophysical specializations that likely contribute to how RA is capable of firing such fast, precise bursts of action potentials.

**Results:**

*Increase of Navβ4 in RA during vocal development*

Recently, we have identified several transcripts for voltage-gated $Na^+$ channel beta (Navβ) auxiliary subunits as either positive (Navβ4) or negative (Navβ3) markers of adult male RA in the arcopallium [104]. While the functional role of these different Navβ subunits in RA neuronal excitability is unknown, Navβ4 is of particular interest because, in combination with Kv3 potassium currents [186], it can promote narrow spike waveforms and high-frequency firing via a resurgent $Na^+$ current ($I_{NaR}$) [187-189]. We, therefore, hypothesized that Navβ4 might be a key

determinant of excitability properties in RA. High Navβ4 expression is restricted to RA in the arcopallium of adult male zebra finches [104], contrasting sharply with Navβ3, which is all but absent in RA (Fig. 4.1A,B). Navβ4 is thought to promote high-frequency firing by limiting classical inactivation of Nav channels and promoting $I_{NaR}$ [189]. Based on its expression levels, we predicted that neurons within RA would have a larger $I_{NaR}$ than those in the arcopallium outside RA. We thus performed whole-cell voltage-clamp (VC) recordings at room temperature in RA sagittal slices from adult male finches (Fig. 4.1C). Voltage steps from −90 to +30 mV elicited large transient $Na^+$ currents ($I_{NaT}$) and, as predicted, subsequent steps to a range of test potentials (+15 to −75 mV) yielded robust $I_{NaR}$ (Fig. 4.1D) with a peak of −3.0 nA ± 0.28 (mean ± SEM) at the −45 mV test potential (Fig. 4.1E). These results reveal an exceptionally large $I_{NaR}$ in RAPNs compared to those recorded in cerebellum [189] and brainstem neurons [190].

In sharp contrast to RAPNs, neurons recorded in a caudal arcopallial region outside RA (Fig. 4.1A; shaded area), which has low Navβ4 expression, showed a much smaller $I_{NaR}$ (Fig. 4.1D). Notably, the recorded cells outside RA were morphologically distinct from neurons within RA, with relatively smaller somata and dendrites with numerous, thin spines. Because $I_{NaR}$ is a function of previously opened Nav channels [189], the expression of which can vary across cells, we normalized peak $I_{NaR}$ measurements to the peak $I_{NaT}$ (ratio of $I_{NaR}/I_{NaT}$). The average normalized peak $I_{NaR}$ was still much larger in RAPNs compared to neurons outside RA over a range of test potentials (−60 to −30 mV), with maximum normalized

values (mean ± SE) of 0.28 ± 0.01 and 0.08 ± 0.01 at −45 mV for RAPNs and neurons outside RA, respectively (Fig. 4.1F).

Male juvenile zebra finches progress through a critical period of vocal learning, during which vocal practice guided by auditory feedback is required to accurately produce a copy of the tutor song [56, 191]. Zebra finches do not sing prior to ~28 days post hatch (dph). During the next phase (~28–45 dph) they produce unstructured vocalizations, referred to as subsong. They next enter a plastic phase (~45–90 dph), during which songs become more structured as the tutee refines syllable structure and sequencing guided by auditory feedback. By ~90 dph the song becomes crystallized and highly stereotyped. These developmental changes in song coincide with the growth of vocal control nuclei in males and changes in the connectivity of RAPNs [20, 42, 55, 88, 91, 192-195]. In contrast, female finches do not develop a song and their song nuclei experience marked reductions in volume, as well as sharp decreases in neuronal cell number and size [42, 88].

To investigate whether developmental changes in RA are also associated with changes in the expression of Navβ subunits, we performed in situ hybridizations for Navβ4 and Navβ3 in sagittal brain sections from male zebra finches at ages known to be within the pre-song (20 dph), subsong (35 dph), plastic song (50 dph), and crystallized song (>90 dph) stages of vocal development [191]. We observed an age-dependent increase in Navβ4 expression levels in RA, with significant differences between 20 and 35 dph juveniles compared to adults (Fig. 4.2A-D black). We also detected an age-dependent increase in the proportion of Navβ4-expressing cells, progressing from 2% of cells relative to Nissl in pre-song juveniles (20 dph) to 24%, 31%,

and 63% in 35 dph, 50 dph, and adult birds, respectively, with a significant difference between adult and 20 dph finches (Fig. 4.2J, black). In contrast, no significant changes in Navβ4 expression levels, or the proportion of positive cells, were detected outside RA across ages (Fig. 4.2A-D,J, gray). In stark contrast to Navβ4, we found that Navβ3 expression levels and the proportion of Navβ3-expressing cells in RA decreased markedly across ages (Fig. 4.2E-H,K,L, black), with little evidence of change outside RA (Fig. 4.2E-H,K,L, gray).

Given the changes in Navβ4 expression, we predicted a corresponding age-dependent increase of $I_{NaR}$ in RAPNs. RA could be readily identified in sagittal brain sections obtained from 20 to 50 dph male finches via infrared differential interference microscopy. Indeed, $I_{NaR}$ was small or absent in RAPNs from 20 dph finches, but its magnitude increased sharply with age (Fig. 4.3A,B), with significant age-dependent effects observed for both peak $I_{NaR}$ (Fig. 4.3C) and normalized peak $I_{NaR}/I_{NaT}$ ratios (Fig. 4.3D). At the −30 mV test pulse, significant differences were found for post hoc pairwise comparisons of raw and normalized $I_{NaR}$ across all age groups. Importantly, the average normalized peak $I_{NaR}/I_{NaT}$values were significantly correlated with both the average Navβ4 expression level (Fig. 4.3E) and the average proportion of Navβ4-expressing cells within RA (Fig. 4.3F). Therefore, Navβ4 expression strongly correlates with $I_{NaR}$ in RA and predicts the increase of $I_{NaR}$ seen in RA across different ages within the vocal learning period.

*Kv3.1 is the TEA-hypersensitive ion channel subunit preferentially expressed in RAPNs*

RAPNs produce remarkably narrow APs (half-width = ~0.2 ms at 40$^{o}$C; [45]), making them well suited for orchestrating rapid movements of the syringeal and respiratory muscles required for song production. Upon examining properties of spontaneous APs, we found that RAPNs and AId neurons shared similar threshold, amplitude, peak and after-hyperpolarization at both temperatures examined (Fig. 4.4C). However, AId neurons had an AP half-width that was twice as broad as the RAPN APs (Fig. 4.4C-D and 4.4F-G). The shorter half-width of RAPNs could be due to either a faster AP depolarization and/or repolarization rate. To determine which phase of the AP was responsible for this difference, we derived phase plane plots from averaged spontaneous APs and compared the maximum rates of depolarization and repolarization. At room temperature, the maximum rate of repolarization was 76% larger in RAPNs compared to AId neurons (Fig. 4.4E), whereas the maximum rate of depolarization was only 29% larger in RAPNs. At 40$^{o}$C the difference in the maximum depolarization rate disappeared, while the maximum repolarization rate was 37% greater in RAPNs (Fig. 4.4H). These results suggest that the relatively slower AP repolarization is predominantly responsible for the broader AP half-width of AId neurons compared to the ultranarrow RAPN APs. We hypothesized that differential usage of voltage gated potassium channels enables the differential repolarization rate between RA and AId.

There are four members of the Shaw-related channel family (Kv3.1-3.4) in vertebrates [196, 197]. We examined each of the genes in the Kv3 family to determine if any exhibit a molecular specialization in RA. As a start, through close assessments of reciprocal alignments and synteny we confirmed that the locus named KCNC1 (100144433; located on chromosome 5) is the zebra

finch ortholog of mammalian KCNC1, noting the conserved synteny across major vertebrate groups (Fig. 4.5A). Importantly, the predicted zebra finch Kv3.1 protein (Kv3.1b isoform) is remarkably conserved (96.48% residue identity) with human (Fig. 4.6) and is thus predicted to have similar pharmacology as in mammals, which is supported by our recordings from RAPNs. Additionally, we confirmed the correct identification of the zebra finch orthologs of mammalian KCNC2/Kv3.2 and KCNC4/Kv3.4, as previously reported [118].

Previous investigations of the zebra finch genome (taeGut1, [29]) reported that the zebra finch lacked a KCNC3/Kv3.3 ortholog [118]. Recent long-read sequencing technology, however, has facilitated a more complete assembly of genomes [32], elucidating the presence of some genes previously thought to be absent. Using the recent RefSeq release 106 (GCF_003957565.2 assembly), we observed a locus (LOC115491734) on the newly assembled zebra finch chromosome 37 described as similar to member 1 of the KCNC family. LOC115491734, however, exhibited the highest alignment scores and conserved upstream synteny with KCNC3/Kv3.3 in humans and various vertebrate lineages, noting that NAPSA in zebra finch and other songbirds is misannotated as cathepsin D-like (LOC121468878) (Fig. 4.5C). We conclude that LOC115491734 is the zebra finch ortholog of mammalian KCNC3, previously thought to be missing in birds [118], and not KCNC1. We note that the downstream immediate synteny is not conserved across vertebrate lineages, with the ancestral condition in tetrapods likely being TBC1D17 downstream of KCNC3. The predicted zebra finch KCNC3 protein showed only moderate conservation with human (68.32% residue identity), some domains including the BTB/POZ and transmembrane domain being fairly conserved, but spans of residues on the N-

terminal and C-terminal regions being highly divergent. Notably, the N-terminal inactivation sequence [197] of KCNC3 appears to be absent in the zebra finch (Fig. 4.7).

We next performed *in situ* hybridization for all identified KCNC/Kv3.x family members in adjacent frontal brain sections from adult male zebra finches. We replicated our previous finding that Kv3.1 expression is higher in RA than in AId ([38]; Fig. 4.8A, top left), whereas expression of both Kv3.2 (Fig. 4.8A, top right) and Kv3.4 (Fig. 4.8A, bottom right) was non-differential between RA and AId. Unlike the graded distribution found in avian [198] and mammalian [199] auditory brainstem, Kv3.1 appeared uniformly distributed across RA. Kv3.2-expressing cells were sparse, strongly labeled, and reminiscent of the GABAergic cell distribution [149], while Kv3.4 expression was uniformly weak throughout both brain regions (Fig. 4.8A). We also found that Kv3.3, while a uniquely specific marker for both brain regions compared to the surrounding arcopallium, is non-differentially expressed between RA and AId (Fig. 4.8A, bottom left). Importantly, we found strong Kv3.3 expression in the Purkinje cell layer of the cerebellum (Fig. 4.8A, bottom left), consistent with findings in mammals [186]. Furthermore, other TEA-hypersensitive potassium channel subunits examined, namely members of the Kv1 (KCNA), BK (KCNMA), and Kv7 (KCNQ) families, had similar expression in RA and AId, with the exception of KCNQ2, which had a lower proportion of labeled cells in RA than in AId (Fig. 4.8B,C). These results point to KCNC1 as the only TEA-hypersensitive subunit more highly expressed in RA compared to AId, providing supporting evidence for a role of Kv3.1 in shaping the ultranarrow AP of RAPNs.

Intriguingly, while curating avian KCNC3s, we discovered a previously undescribed KCNC family member in the genomes of several bird species, but notably absent in songbirds (Fig. 4.5B). This gene most closely resembled KCNC1 in predicted domains and amino acid conservation, thus we named this KCNC1 paralog as KCNC1L. We also observed KCNC1L in non-avian sauropsids including lizards and snakes, where the locus seems to be duplicated. It was not present in humans/mammals, nor in amphibian or fish outgroups. This suggests this paralog possibly arose after the split between mammals and sauropsids, with a subsequent loss in songbirds. While helping to further solidify the differential expression of KCNC1/Kv3.1 as key for RAPN physiology, these findings bring new insights into the evolution of this important family of neuronal excitability regulators in vertebrates. How the newly identified avian KCNC3 and non-oscine KCNC1L contribute to avian neuronal physiology are intriguing questions for further study.

## Discussion

Comparative studies in the auditory system have suggested convergent strategies for temporal coding of sound stimuli in birds and mammals [200, 201]. We have recently described how upper-motor RAPNs and L5PNs in mammalian M1 also share convergence in factors determining the AP initiation and upstroke. This includes APs with biphasic depolarization rates, persistent $Na^+$ currents, large transient $Na^+$ currents ($I_{NaT}$) with rapid kinetics, and high Navβ4 mRNA expression, which we showed to be linked to robust resurgent currents ($I_{NaR}$) in RAPNs [45]. A major difference, however, was the ultranarrow AP waveform of RAPNs that is also a

unique trait of the large Betz cells found in Layer 5 of M1 in cats [202] and primates [168]. Here we show RAPNs produce narrower APs compared to AId neurons, due largely to higher maximum repolarization rates. In contrast to RAPNs, the 1.2 ms AP half-width at 24$^{o}$C and the "regular" non-adapting AP firing of AId neurons render them more similar to canonical L5PNs in the motor cortex of rodents [203], cats [202] and primates [168]. Thus, like Betz cells [146, 169, 204], RAPNs appear to be a specialized class of upper motor neurons that display higher temporal precision of AP firing and faster firing rates than AId neurons.

We also note striking differences between the properties of finch RAPNs and AId neurons compared to those of L5PNs in mammalian M1. Foremost, RAPNs and AId neurons lack the large, tufted apical dendrites typical of L5PNs [205]. Additionally, these avian neurons fire spontaneously in the absence of synaptic inputs, a property not typically seen in M1 L5PNs [202, 203]. The size and distribution of Betz cells in M1 also differ compared to RAPNs in the finch. Whereas Betz cells have very large somas and are interspersed with smaller L5PNs [206, 207], RAPNs and AId neurons have similar soma sizes and localize to adjacent but distinct regions within the finch arcopallium [38, 82, 83].

In conclusion, RAPNs in zebra finches exhibit many fundamental molecular and functional similarities to primate Betz cells, that may be involved is fine digit movements [208]. In combination with their well-defined role in singing behaviors, this study identifies RAPNs as a novel and more accessible model for studying the properties of Betz-like pyramidal neurons that offer the temporal precision required to control complex learned motor behaviors.

## Methods

Animal subjects:

All of the work described in this study was approved OHSU's Institutional Animal Care and Use Committee (Protocol #: IP0000146) and is in accordance with NIH guidelines. Zebra finches (*Taeniopygia guttata*) were obtained from our own breeding colony. All birds used were male and > 120 days post hatch. Birds were sacrificed by decapitation and their brains removed. For electrophysiology experiments brains were bisected along the midline, immersed in ice-cold cutting solution, and processed as described below. For *in situ* hybridization experiments brains were cut anterior to the tectum and placed in a plastic mold, covered with ice-cold Tissue-Tek OCT (Sakura-Finetek; Torrance, CA), and frozen in a dry ice/isopropanol slurry and processed as described below.

In situ hybridization:

To compare mRNA expression levels for *SCN3B, SCN4B, KCNC1, KCNC2, KCNC3, KCNC4, KCNMA1, KCNQ2, KCNQ3, KCNA1, KCNA2, and KCNA6* across RA and AId, brains sections (thickness = 10 µm) were cut coronally on a cryostat and mounted onto glass microscope slides (Superfrost plus; Fisher Scientific, Hampton, NH, USA), briefly fixed, and stored at -80°C. For each brain, every $10^{th}$ slide was fixed and stained for Nissl using an established cresyl violet protocol. Slides were examined under a bright-field microscope to identify sections containing the core region of RA and AId as previously defined [38]. *In situ* hybridization was conducted

using an established protocol [21]. Briefly, slides were hybridized under pre-optimized conditions with DIG-labeled riboprobes synthesized from BSSHII-digested cDNA clones obtained from the ESTIMA: Songbird clone collection [209]. Specific clones corresponded to GenBank IDs FE734016 (SCN3B; NavB3), FE730991 (SCN4B; NavB4), CK302978 (KCNC1; Kv3.1), DV951094 (KCNC2; Kv3.2), DV953393 (KCNC3; Kv3.3), CK308792 (KCNC4; Kv3.4), DV954467 (KCNMA1; BK), FE737967 (KCNA1; Kv1.1), FE720882 (KCNA2; Kv1.2), FE733881 (KCNA6; Kv1.6), DV954380 (KCNQ2; Kv7.2), and CK316820 (KCNQ3, Kv7.2). After overnight hybridization, slides were washed, blocked, incubated with alkaline phosphatase conjugated anti-DIG antibody (1:600; Roche, Basal, Switzerland) and developed overnight in BCIP/NBT chromogen (Perkin Elmer; Waltham, MA, USA). Slides were coverslipped with VectaMount (Vector, Newark, CA, USA) permanent mounting medium, and then digitally photographed at 10X under bright field illumination with a Lumina HR camera mounted on a Nikon E600 microscope using standardized filter and camera settings. Images were stored as TIFF files and analyzed further using the FIJI distribution of ImageJ [210]. We note that high-resolution parasagittal images depicting expression of *KCNC1*, *KCNC2, KCNA1*, *KCNA6, KCNQ2,* and *KCNMA1* in RA of adult male zebra finches are available on the Zebra Finch Expression Brain Expression Atlas (ZEBrA; www.zebrafinchatlas.org). All probes were evaluated for specificity by examining their alignment to the zebra finch genome and avoiding probes with significant cross alignments to other loci (as detailed previously [22]).

For each gene, we quantified both expression levels based on labeling intensity (i.e. average pixel intensity) and the number of cells expressing mRNA per unit area. We measured the

average pixel intensity (scale: 0-256) in a 200 x 200 µm window placed over each target area in the images of hybridized sections. To normalize signal from background we subtracted an average background level measured over an adjacent control area in the intermediate arcopallium that was deemed to have no mRNA expression. The expression ratio was calculated as $RA_{OD}/AId_{OD}$ where values greater than 1 are more highly expressed in RA and values less than 1 are more highly expressed in AId. We also quantified the number of labeled cells in each arcopallial region by first establishing a threshold of expression 2.5X above the background level. Standard binary filters were applied and the FIJI 'Analyze Particles' algorithm was used to count the number of labeled cells per 200 $\mu m^2$.

Slice preparation for electrophysiology experiments:

Frontal (180 µm for current clamp and 150 µm for voltage clamp) slices were cut on a vibratome slicer (VT1000, Leica) in an ice-cold cutting solution containing (in mM): 119 NaCl, 2.5 KCl, 8 $MgSO_4$, 16.2 $NaHCO_3$, 10 HEPES, 1 $NaH_2PO_4$, 0.5 $CaCl_2$, 11 D-Glucose, 35 Sucrose pH 7.3-7.4 when bubbled with carbogen (95% $O_2$, 5% $CO_2$; osmolarity ~330-340 mOsm). Slices were then transferred to an incubation chamber containing artificial cerebral spinal fluid (aCSF) with (in mM): 119 NaCl, 2.5 KCl, 1.3 $MgSO_4$, 26.2 $NaHCO_3$, 1 $NaH_2PO_4$, 1.5 $CaCl_2$, 11 D-Glucose, 35 Sucrose pH 7.3-7.4 when bubbled with carbogen (95% $O_2$, 5% $CO_2$; osmolarity ~330-340 mOsm) for 10 min at 37°C, followed by a room temperature incubation for ~30 min prior to start of electrophysiology experiments.

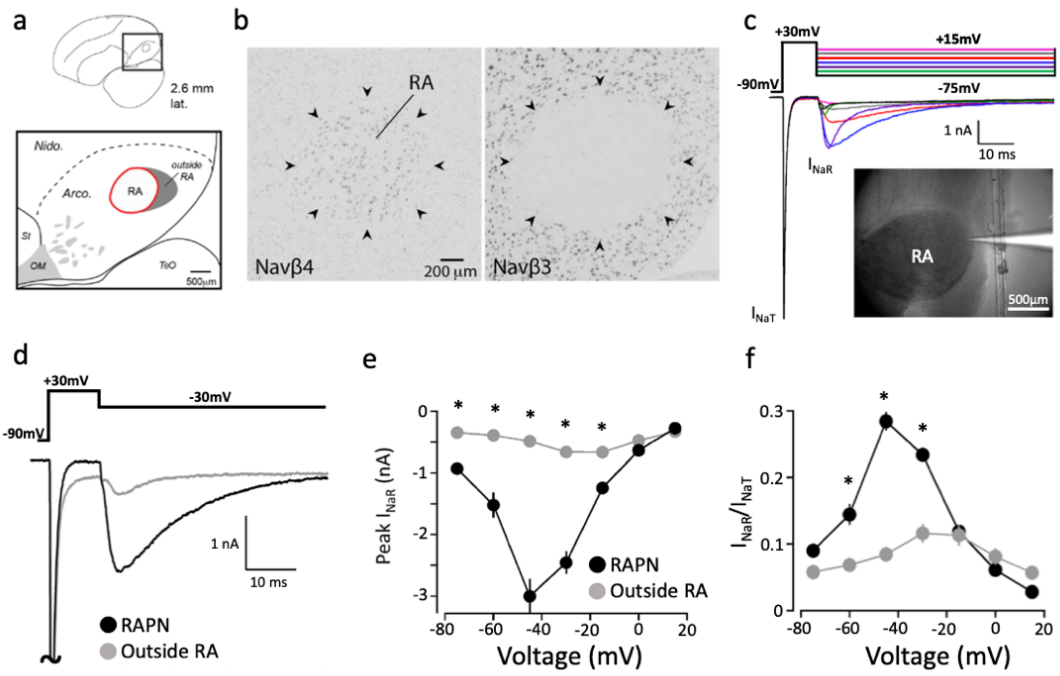



Fig. 4.1. Expression of Navβ4 and β3 mRNAs, resurgent $Na^+$ current ($I_{NaR}$), and the spiking properties of RAPNs in the arcopallium of adult male zebra finches. A. Anatomical drawing indicating the position of RA within the arcopallium in a parasagittal section; distance from the midline indicated; the shaded area represents the region outside RA that is compared with recordings within RA. Arco., arcopallium; Nido., Nidopallium; OM, occipitomesencephalic tract; St, striatum; TeO, Optic tectum. Orientation: dorsal is up and anterior to the left. B. In situ hybridization images show high Navβ4 expression (left) and lack of Navβ3 expression (right) in RA (indicated by arrowheads). C. Representative examples of $I_{NaT}$ and $I_{NaR}$ elicited in an RAPN at 24°C by the voltage clamp protocol at the top. $I_{NaT}$ was elicited via a 10 ms step to +30 mV followed by test potentials to +15 (pink), 0 (gray), −15 (red), −30 (blue), −45 (purple), −60 (green), and −75 mV (black) to elicit $I_{NaR}$. The cell was held at −90 mV during a 2 s intersweep interval. Inset: Image of a patch pipette filled with fluorescent dye for RA recordings. D. Detail views of example $I_{NaT}$ and $I_{NaR}$ elicited in an RAPN (black) and in a neuron outside RA (gray) by the voltage clamp protocol at the top. The $I_{NaT}$ peaks have been truncated. E. Average I–V curves for the peak $I_{NaR}$ in RAPNs (black) and in neurons outside RA (gray) (two-way ANOVA with Tukey's post hoc; $P = 1.7 \times 10^{-21}$, $F(6, 146) = 27.00$, N = 12 RAPNs and 11 neurons outside RA; individual comparisons: −15 mV ($P = 0.009$), −30 mV ($P = 1.7 \times 10^{-18}$), −45 mV ($P = 3.0 \times 10^{-29}$), −60 mV ($P = 1.3 \times 10^{-8}$), −75 mV ($P = 0.008$)). Data are presented as mean values ± SEM. F. Average I–V curves after normalization of the peak $I_{NaR}$ to the peak $I_{NaT}$ measured in a given sweep then averaged across cells (two-way ANOVA with Tukey's post hoc; $P = 6.1 \times 10^{-23}$, $F(6, 147) = 29.36$, N = 12 RAPNs and 11 neurons outside RA; Individual comparisons: −30 mV ($P = 1.1 \times 10^{-11}$), −45 mV ($P = 1.8 \times 10^{-26}$), −60 mV ($P = 1.3 \times 10^{-5}$)). Data are presented as mean values ± SEM.

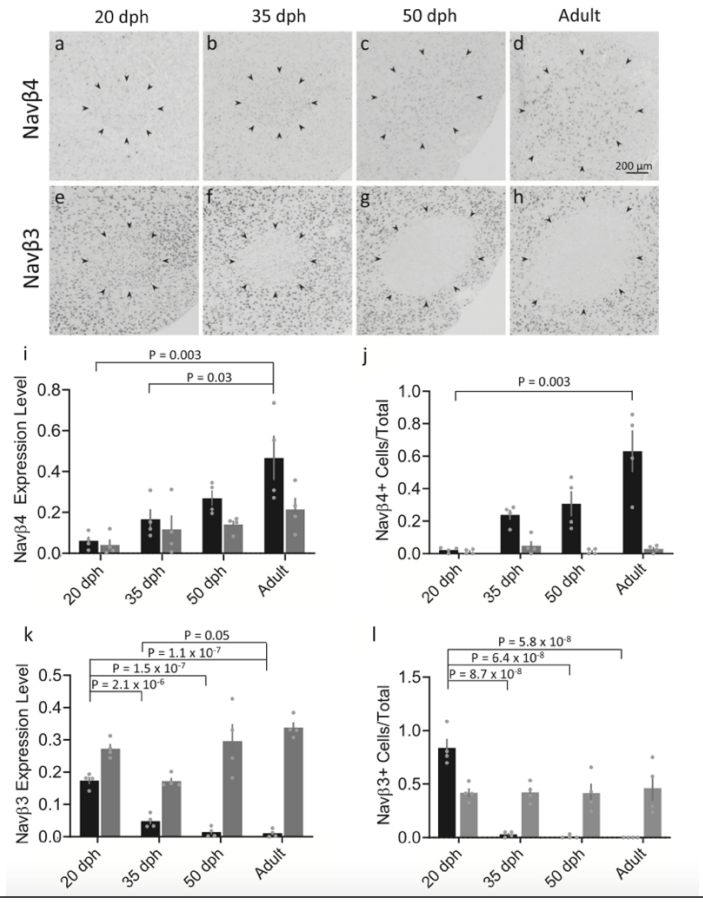



Fig. 4.2. Age-dependent changes in expression of Navβ3 and Navβ4 mRNAs in the arcopallium of male zebra finches. a–d Representative in situ hybridization images for Navβ4 mRNA within RA across ages indicated by days post hatch (dph); black arrowheads indicate RA borders. e–h. Representative in situ hybridization images for Navβ3 mRNA within RA across ages. Black arrowheads indicate RA borders. i Comparison of Navβ4 expression levels (normalized optical density) across age groups within RA (black) and in a caudal arcopallial region of equal size outside RA (gray). Significant age differences were observed in RA (one-way ANOVA with Tukey's post hoc; P = 0.005, F (3, 12) = 7.416, N = 4 males per age). Data are presented as individual data points with bars as mean values ± SEM. j Comparison of the proportions of Navβ4-expressing cells across age groups within RA (black) and in an arcopallial region of equal size outside RA (gray). Significant age differences in RA (one-way ANOVA with Tukey's post hoc; P = 0.001, F = F (3, 12) = 10.79, N = 4 males per group). Data are presented as individual data points with bars as mean values ± SEM. k Comparison of Navβ3 expression level (normalized optical density) across age groups within RA (black) and in an arcopallial region of equal size outside RA (gray). Significant age differences in RA (one-way ANOVA with Tukey's post hoc; $P = 5.4 \times 10^{-8}$, F (3, 12) = 73.43, N = 4 males per group). Data are presented as individual data points with bars as mean values ± SEM. l Comparison of the proportions of Navβ3-expressing cells across age groups within RA (black) and in an arcopallial region of equal size outside RA (gray). Significant age differences in RA (two-way ANOVA with Tukey's post hoc; $P = 1.6 \times 10^{-8}$, F (3, 12) = 90.96, N = 4 males per group). Data are presented as individual data points with bars as mean values ± SEM.

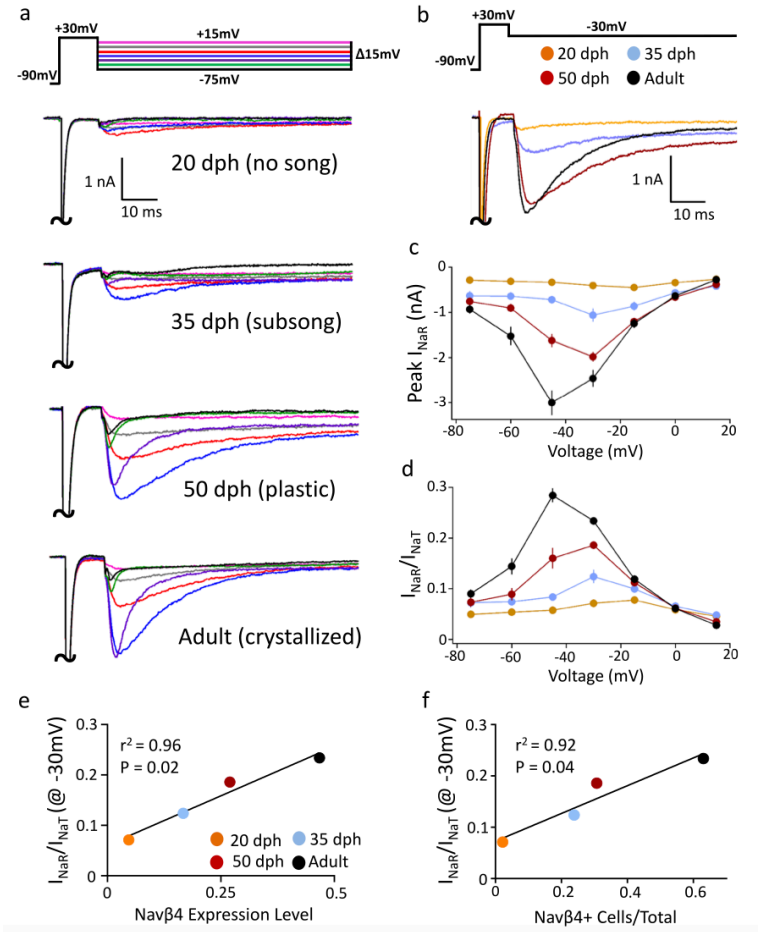



Fig. 4.3. Age-dependent changes in $I_{NaR}$ in RAPNs of male zebra finches. A. Examples of transient ($I_{NaT}$) and resurgent ($I_{NaR}$) $Na^+$ currents elicited across ages (days post hatch, dph) by the voltage clamp protocol shown at the top. The large $I_{NaT}$ peaks have been truncated. B. Representative currents from RAPNs at each age group during the −30 mV test potential shown at the top. The large $I_{NaT}$ peaks have been truncated. c Average I–V curves for the peak $I_{NaR}$ in RAPNs from each age group (two-way ANOVA with Tukey's post hoc comparisons; $P = 3.7 \times 10^{-35}$, F (18, 280) = 16.98, N (cells/age) = 10/20 dph, 12/35 dph, 10/50 dph and 12/adults). Data are presented as mean values ± SEM. d Average I–V curves after normalization of the peak $I_{NaR}$ to the peak $I_{NaT}$ measured in a given sweep then averaged across cells (two-way ANOVA with Tukey's post hoc comparisons; $P = 4.2 \times 10^{-46}$, F (18, 280) = 23.72, N (cells/age) = 10/20 dph, 12/35 dph, 10/50 dph, and 12/adults). Data are presented as mean values ± SEM. e Linear regression between the normalized average peak $I_{NaR}$ values at the −30 mV test potential across ages and the average Navβ4 expression level in RA (two-tailed; no adjustment for multiple comparisons). f Linear regression between the normalized average peak $I_{NaR}$ values at the −30 mV test potential across ages and the proportion of Navβ4-expressing cells in RA (two-tailed; no adjustment for multiple comparisons).

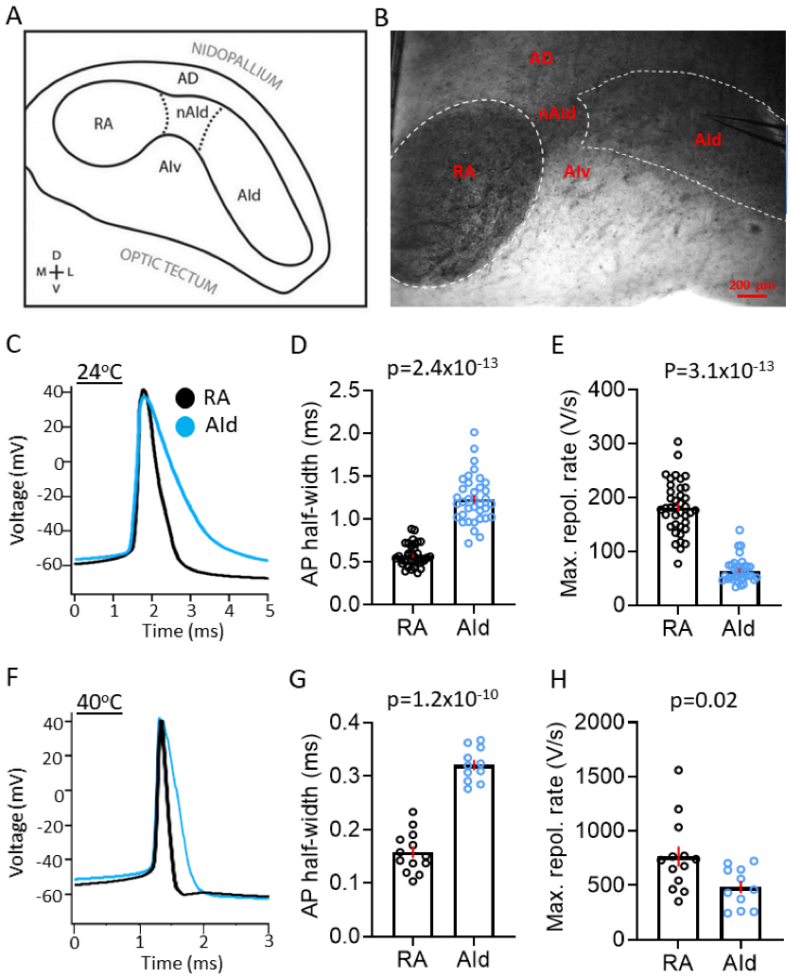



Figure 4.4. Spontaneous AP properties in RAPNs and AId neurons from adult male zebra finches at room and physiological temperatures. A. Drawing of the zebra finch arcopallium in the frontal plane, depicting RA and AId as defined by SCN3B staining in adult males (Nevue et al., 2020). B. IR-DIC image of a frontal brain slice from which RA and AId are clearly visible. Labels correspond to those depicted in (A). Note the recording electrode in AId. C. Overlay of averaged spontaneous APs recorded at ~24°C from a RAPN (black) and AId neurons (blue). D. Comparison of the average spontaneous AP half-widths from RAPNs and AId neurons recorded at ~24°C (Mann-Whitney *U*= 11, two-tailed, N= 39 RAPNs and 36 AId neurons. Red bars indicate standard error). E. Comparison of the average spontaneous AP maximum repolarization rate from RAPNs and AId neurons recorded at ~24°C. Mann-Whitney *U*= 14, two-tailed N= 39 RAPNs and 36 AId neurons. Red bars indicate standard error. F. Overlay of averaged spontaneous APs recorded at ~40°C from a RAPN (black) and AId neurons (blue) respectively. G. Comparison of the average spontaneous AP half-widths from RAPNs and AId neurons recorded at ~40°C. Student's t-test, two-tailed, $t_{stat}$ = 11.33, N= 13 RAPNs and 11 AId neurons. Red bars indicate standard error. H. Comparison of the average spontaneous AP maximum repolarization rate from RAPNs and AId neurons recorded at ~40°C.

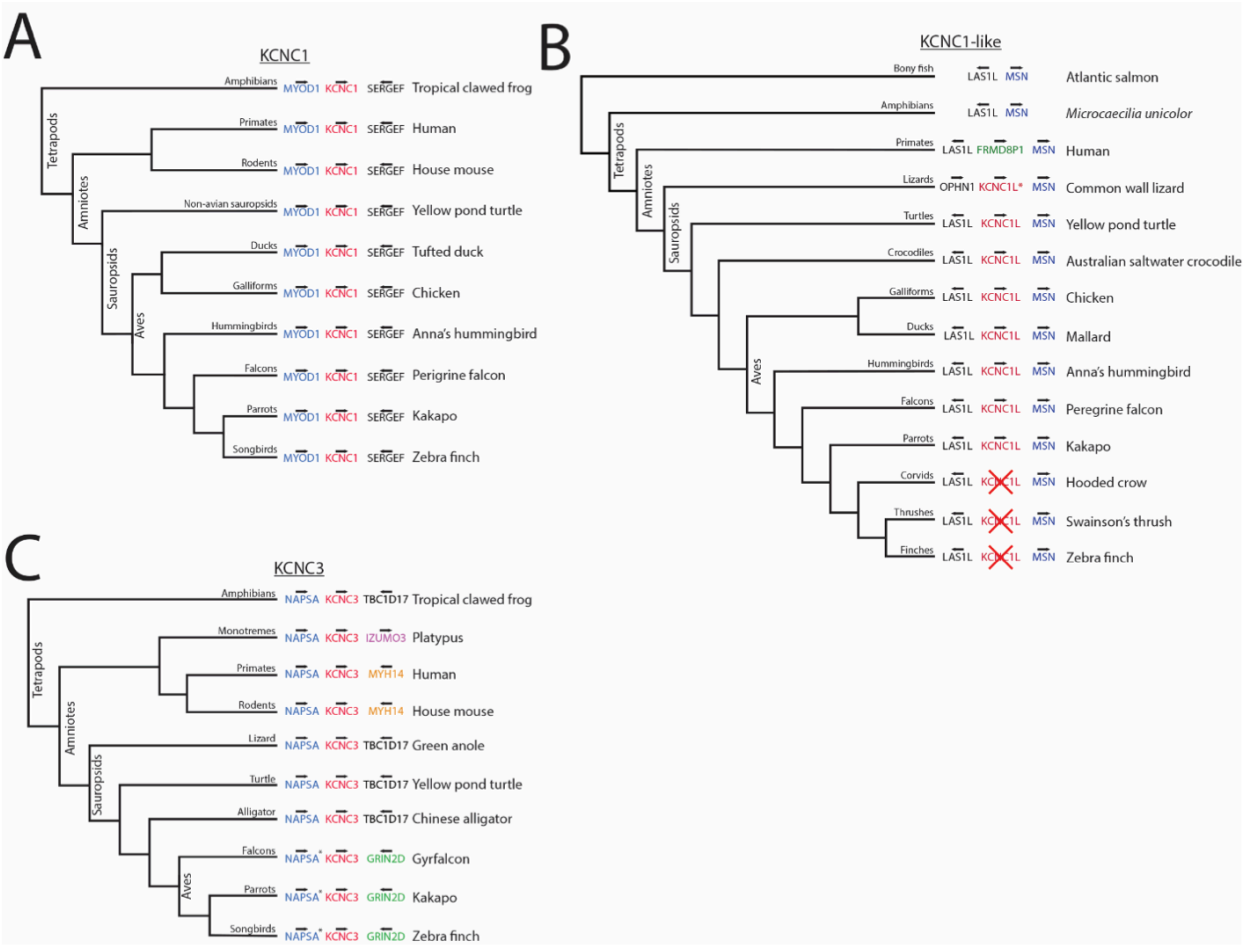



Figure 4.5. Comparative genomics of KCNC/Kv3 family members. A. Simplified cladogram for KCNC1 showing conserved synteny throughout vertebrate groups. B. Simplified cladogram for KCNC1-like paralog present in non-oscine sauropsids. Asterisk denotes a possible duplication at the KCNC1L locus in lizards. Red 'X' denotes gene is not present. C. Simplified cladogram for KCNC3 showing the presence in birds with partial conserved synteny across vertebrate groups. Asterisk denotes NAPSA annotated as cathepsin D-like in birds. Note: For all cladograms, branch lengths are arbitrary and not calibrated for time, and only selected extant branches from major vertebrate groups are shown, to illustrate synteny conservation/divergence.

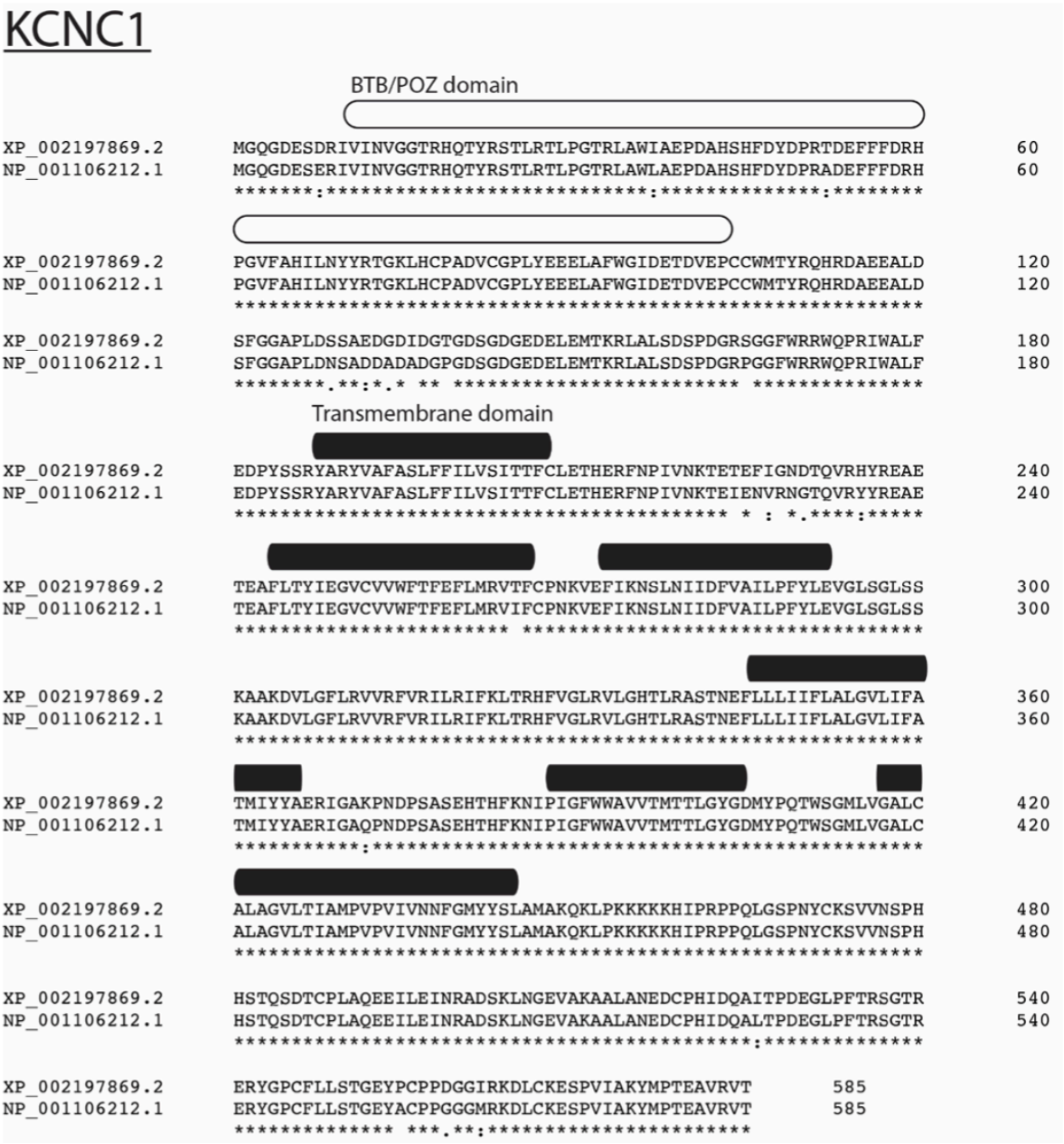



Figure 4.6. Amino acid alignment between Kv3.1 in human (bottom) and zebra finch (top). The two orthologs exhibit a high degree of amino acid conservation.

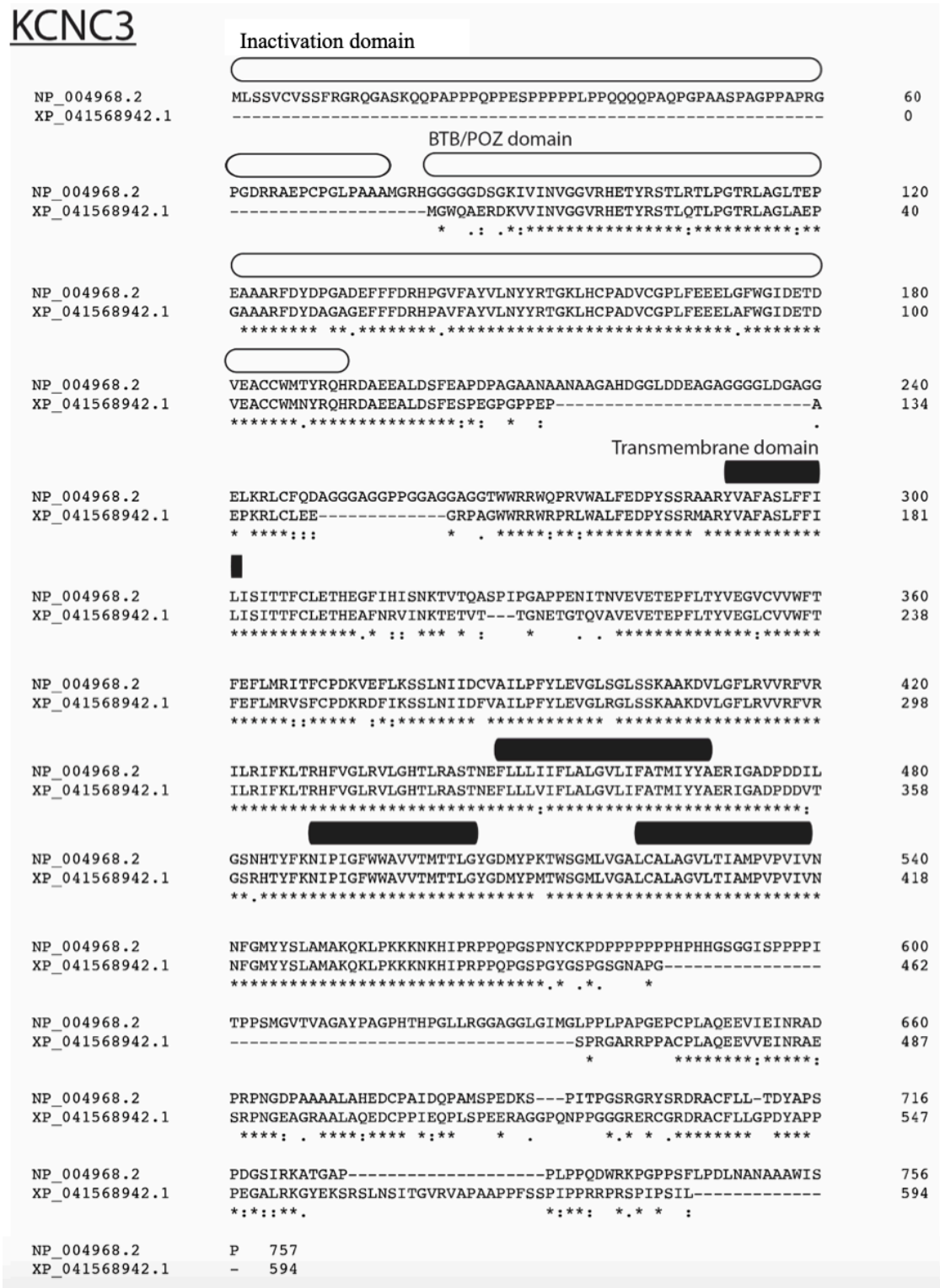



Figure 4.7. Amino acid alignment between Kv3.3 in human (top) and zebra finch (bottom). Amino acid conservation is moderate and zebra finch appears to lack N-terminal domain present in humans.

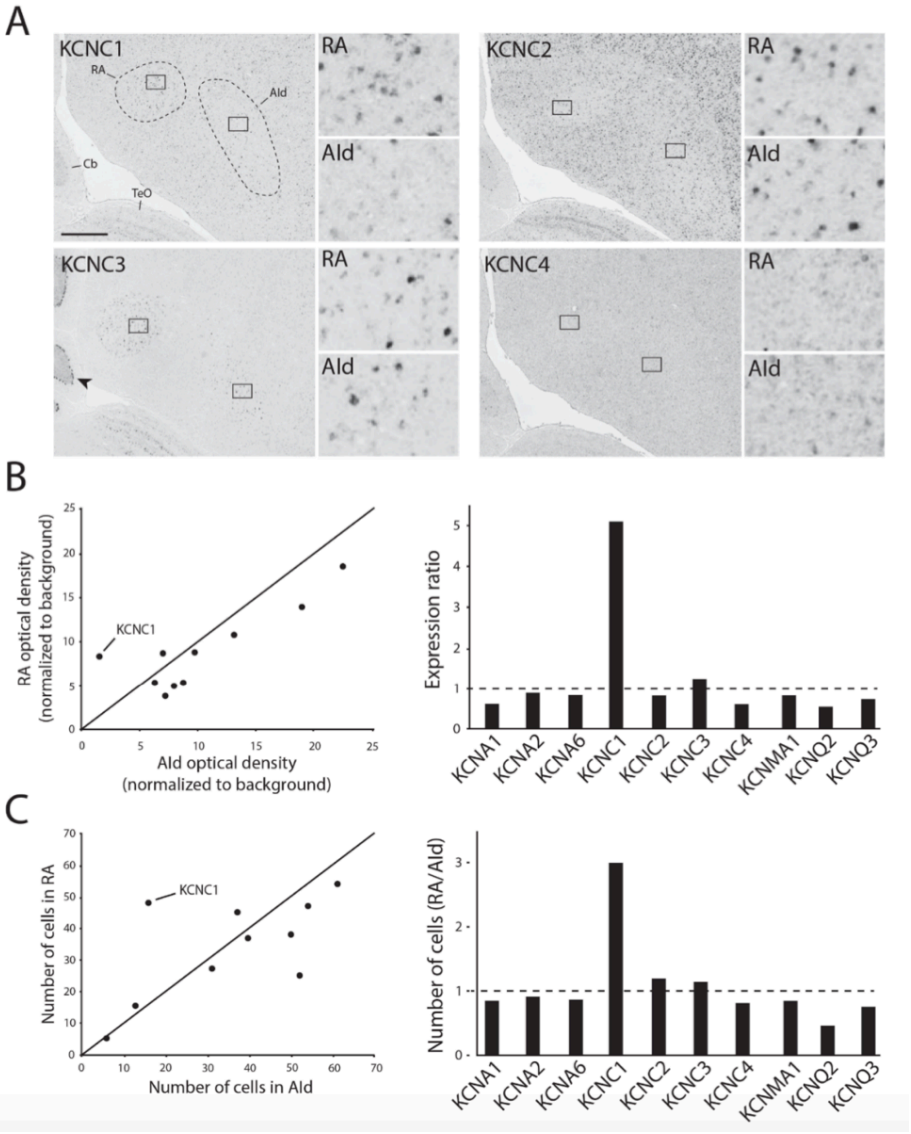



Figure 4.8. $K^+$ channel diversity in the zebra finch arcopallium: Stronger expression of KCNC1 (Kv3.1 subunit) transcripts in RA than in AId. A. Representative *in situ* hybridization images for Kv3 channel family member transcripts in RA (left) and AId (right), from nearly adjacent frontal sections of adult males. Squares in large images depict position of counting windows and of inset images for RA and AId. Arrow points to strong Kv3.3 mRNA staining in the Purkinje cell layer in the cerebellum. *Cb*- cerebellum, *TeO*- Optic tectum. Scale bar: 500 µm. B. Optical density measurements (background subtracted) in RA and AId for subunits associated with TEA-hypersensitive Kv channel types. Expression ratio (right) was calculated as RAOD/AIdOD. C. Labeled cell counts in RA and AId for subunits associated with TEA-hypersensitive Kv channel types. Cell count ratios of RA/AId are shown on right.

# Chapter 5

## Naturally occurring gene knockout attenuates AAV transduction in songbirds

Alexander A. Nevue, Samuel J. Huang, Anusha Sairavi, Hiroyuki Nakai, Claudio V. Mello. AAN, HN, and CVM conceived the study. AAN performed genomics analysis. AAN and SJH performed BC-Seq experiments. AAN and AS performed cell culture experiments. AAN and CVM wrote the text.

**Abstract**

Adeno-associated virus (AAV) is becoming the chosen vector for gene delivery in basic science research as well as for clinical development for gene therapy. Despite this development, how AAVs transduce and deliver gene cargo to a cell is largely unknown. Recently, the genes KIAA1319L (also known as AAVR) and GPR108 were identified from to be critical for AAV transduction. Here we show that GPR108 is absent in the genome of most sauropsids including all birds. We rescue susceptibility to AAV transduction by introducing human GPR108 to avian cells *in vitro*. Furthermore, we find that glycan binding is critical for robust transduction. Then, using a barcoded capsid strategy, we found that AAV9 has the highest transduction efficiency out of wild type serotypes in the songbird brain. These results enhance the understanding of the barriers to AAV gene delivery which is critically important for the design of targeted of therapeutics and provides insight into the evolutionary mechanisms of virus-host interactions.

## Introduction

Adeno-associated virus (AAV) is a non-pathogenic virus that is being developed as a vector for a variety of gene therapy applications. While the FDA has approved two AAV-based gene therapies, with more in clinical development, how AAV enters and delivers its cargo to a cell is not well understood. Understanding vertebrate species variations in the genes involved with AAV transduction may provide novel insights into AAV transduction mechanisms, which may then be leveraged for more efficient gene therapy and gene manipulation applications.

Genome-wide screens have identified the adeno-associated virus receptor (AAVR; also known as *KIAA0319L*) as the near-universal transmembrane receptor protein for AAV entry [52], with *GPR108* and *TM9SF2* subsequently identified as also critical for AAV transduction [53, 211]. However, the transduction efficiencies of the various known AAV serotypes (capsid variants) vary widely across tissues, cell types, and species, and the exact roles of AAV entry factors and their interactions with serotype variants are unclear. To better understand these interactions between AAV serotypes and entry factors across cell types and tissues, we investigated AAV transduction in songbirds, where reports of AAV transduction have been varied but generally low [50, 212, 213].

## Results

We first investigated the zebra finch genome [32] for AAV entry factors. Zebra finch AAVR had only moderate conservation (67.75% amino acid identity) with the mammalian ortholog. The

PKD domains, including PKD2 which is critical for AAV9 [52], showed higher conservation between zebra finch and human (80-92% amino acid conservation), suggesting the domains that are key for transduction may be present in zebra finch AAVR (Fig 5.1A). Strikingly, GPR108, the second most highly enriched gene in the AAVR screen [52], was absent in zebra finch, as well as in all other avian genomes available in NCBI, including the basal chicken, suggesting that this gene is absent in all birds. GPR108 was also absent in the genomes of most non-avian sauropsids such as alligators, snakes, and lizards (Fig. 5.1B), although intriguingly it was present in the turtle lineage, suggesting it may have been lost independently in different lineages rather than in a common ancestor to all sauropsids. Notably, a cluster of mammalian genes syntenic to GPR108 was also missing in birds, suggesting this location of the genome is a hotspot for gene loss (Fig. 5.1B). All other genes identified as involved in AAV transduction in mammals were present in the zebra finch genome, and had similar predicted peptide length and sequence identity with their mammalian orthologs (Fig. 5.1C). The top genes from AAV transduction screens in mammalian cells, including AAVR and TM9SF2 which have been independently verified as critical for nearly all AAV serotype transduction [211], showed broad expression throughout the brain of the zebra finch (Fig. 5.1D).

To determine if the absence of GPR108 impacts the efficiency of AAV transduction, we used a zebra finch cell line derived from embryonic fibroblasts [214] and generated transgenic zebra finch cells that express human GPR108 (hGPR108). Compared to wild type cells, hGPR108+ cells were more susceptible to transduction with AAV2 and AAV9, which are GPR108-dependent in mammals, but not with AAV5, which is GPR108 independent (Fig. 5.1E,F), consistent with a key

role of GPR108 in AAV transduction in finch cells. As many cells were still not transduced, we next tested the chimera serotype AAV2G9, which has galactose binding domains inserted into the AAV2 serotype capsid. AAV2G9 transduced hGPR108+ cells more robustly than AAV2 (Fig. 5.1E, right), indicating that robust AAV transduction is dependent on both the presence of GPR108 and the capacity to bind to glycans on the surface of zebra finch cells (Fig. 5.1G).

Songbirds have a dedicated brain circuitry for singing, and are the choice model organism to study song learning, analogous to human speech acquisition. While there is a pressing need to identify effective viral vectors for assessing the molecular genetics of this rare behavior, a systematic assessment of transduction efficiency of common AAV serotypes has yet to be performed. Comparing different AAV preparations poses multiple confounds as factors such as titer, promoter, and source can vastly impact the efficiency of those viruses. To overcome these factors, we utilized an AAV library containing DNA barcoded capsids under a ubiquitous U6 promoter, which allowed us to quantify the transduction efficiency of AAV1-11 serotypes using Illumina sequencing in an unbiased, high-throughput manner [51]. We injected the library into the arcopallium and striatum of an adult male zebra finch (Fig. 5.2A, inset), and the sequencing reads from extracted RNA were quantified and normalized to the source library producing a phenotypic difference (PD) value which is a measurement of AAV transduction. Overwhelmingly, AAV9 had the highest transduction efficiency in both brain regions (Fig. 5.2A), possibly due to the galactose binding domains of AAV9. To directly examine AAV9 transduction efficiency in the songbird brain, AAV9 under the control of the neuronal specific promoter hSynI or the ubiquitous promoter CMV was injected into the arcopallium.

AAV9 was able to transduce a moderate number of cells in vivo with the hSynI promoter and even more with the CMV promoter. We also compared AAV9 transduction with VSV G lentivirus and Ad5 adenovirus, both commonly used for gene transfer applications. Both lentivirus and adenovirus were inefficient at transducing zebra finch neurons. Lentivirus-labeled cells were sparse but GFP signal was brighter and filled the proximities of the neuron better than adenovirus (Fig. 5.2B). Low transduction with VSV G lentivirus is in agreement with a previous study showing that the VSV G receptor (LDLR) is truncated and inactive in zebra finches [215]. Given the low transduction with Ad5 adenovirus, entry factors for adenovirus may also be lacking in the zebra finch genome. We then tested a chimera capsid that is comprised of an AAV9 backbone with the VP1u region of AAV5 that has been shown to confer GPR108 independence (Dudek et al., 2020) (Fig. 5.2C). We found that despite the genomic lack of GPR108, the chimera capsid (AAV5.9) was unable to increase transduction in the songbird brain (Fig. 5.2D). This suggests that the VP1 region of AAV9 is critical to transduction in vivo, likely for cellular entry. In summary, AAV9 produces high transduction in the zebra finch brain and is a useful tool for future gene manipulation studies.

## Discussion

Our findings show that the genomic lack of GPR108 is associated with low AAV transduction in avian cells, and the reintroduction of human GPR108 is sufficient for transduction. The lack of GPR108 extends to other bird groups and most non-avian sauropsids and may have been the result of a genomic loss in an archosaur ancestor. Strains of AAVs have been isolated from

these species lacking GPR108 suggesting other mechanisms have evolved to not be dependent on GPR108, such as with AAV5 [53]. In tandem with our previous report [215], the songbird genome has multiple losses of functional viral receptors suggesting the genomic landscape for host-viral interactions is complex which is relevant for emerging zoonotic viruses.

The data presented here shed further light on the importance of deep transcriptional and biochemical investigation of target cells for AAV mediated gene therapy. The brain, for example, is a complex tissue containing many highly distinct regions each with a multitude of neuronal and non-neuronal cell types. With many therapeutics relying on efficient delivery to specific cell types, and the probable diversity in cell surface attachment factors and expression levels of entry factor proteins, development of engineered AAVs for gene therapy should be done with these transduction components in consideration.

## Methods

### *Genomics*

We searched for avian orthologs of GPR108 in four songbirds (*Taeniopygia guttata*, *Catharus ustulatus*, *Corvus moneduloides*, *Pseudopodoces humilis*), a pigeon (*Columba livia*), a duck (*Anas platyrhynchos*), a falcon (*Falco rusticolus*), and a chicken (*Gallus gallus*). These avian species were chosen because they contained the flanking genes C3 and VAV1 with no intervening sequence gaps noting that this region of the genome is within microchromosome 30, which has been very difficult to fully sequence and assemble in birds, typically requiring long-read Pacbio

technology [32]. The presence or absence of GPR108 was assessed by combining curation of current annotations in Refseq with BLAST alignments with identified orthologs in non-avian species, as well as comparative synteny analysis. Besides mammals (*Homo sapiens*, *Mus musculus*), GPR108 was found in an amphibian (*Microcaecilia unicolor*) and a turtle (*Chelonia mydas*), but not found in a lizard (*Zootoca vivipara*), a snake (*Pantherophis guttatus*), and an alligator (*Alligator mississippiensis*). For the AAVR analysis, the sequences of KIAA0319L from zebra finch (Gene ID: 100217975) and human (Gene ID: 79932) were imported into InterProScan for domain prediction and MUSCLE was used for sequence alignments.

*In situ hybridization*

*In situ* hybridizations for *AAVR* (CK304065), *TM9SF2* (FE735586), *VPS29* (CK304285), *JTB* (FE721652), *SLC35B2* (CK316607), and *ATP2C1* (DV948245) were carried out as previously described [21, 38]. Briefly, plasmids containing cDNA of the gene of interest were isolated and restriction enzyme digested with BSSHII to release the insert. The insert was purified using a QIAquick PCR purification kit (Qiagen). Antisense DIG-tagged probes were synthesized by *in vitro* transcription using T3 polymerase (Promega) and purified using a Sephadex G-50 column. Probes were hybridized to sections overnight at 65°C and washed. Sections were blocked an incubated in anti-DIG-AP Fab fragments (1:600; Roche) for two hours and incubated in BCIP/NBT chromogen (PerkinElmer) overnight. Probe specificity was examined by aligning the corresponding sequences to the zebra finch genome and verifying that alignment was specific to the expected loci.

*Cell Culture*

The cell line CFS414 derived from zebra finch embryonic fibroblasts was acquired from the Jarvis lab [214]. Cells were cultured in DMEM supplemented with 10% FBS and 1% each of L-Glutamine and Penicillin-Streptomycin. To generate the transgenic line expression hGPR108, 1.9 µg of the transposon (pSBbi-RP, Addgene plasmid #60513) [216] and 100 ng of the transposase (pCMV(CAT)T7-SB100, Addgene plasmid #34879) [217] were transfected using polyethylenimine (PEI) with a ratio DNA:PEI of 1:2. One day after transfection, the cells were subjected to selection with 1 µg/mL puromycin and selection continued for 5 days until all cells were RFP+. Human GPR108 (NCBI RefSeq NP_001073921) was synthesized as GeneArt Strings by Invitrogen and cloned into pSBbi using Gibson cloning.

*Viral production*

AAVs were prepared as previously described [51]. Cloning of the VP1u region of AAV5 was done as previously described [53].

HDAd5-hSyn-eGFP adenovirus was acquired from the University of Iowa Viral Vector Core with a titer of $6.30x10^{10}$ IGU/mL. VSV G pseudotyped lentivirus was prepared at the OHSU virology core using pHR-hSyn-eGFP plasmid (Addgene plasmid #114215) with a titer of $1.02x10^{6}$ TU/uL.

*Stereotactic injections and tissue processing*

All procedures involving live animals were approved by the Institutional Animal Care and Use Committee at Oregon Health & Science University and in accordance with NIH guidelines. Adult male zebra finches (*Taeniopygia guttata*) were obtained from our colony or purchased from a

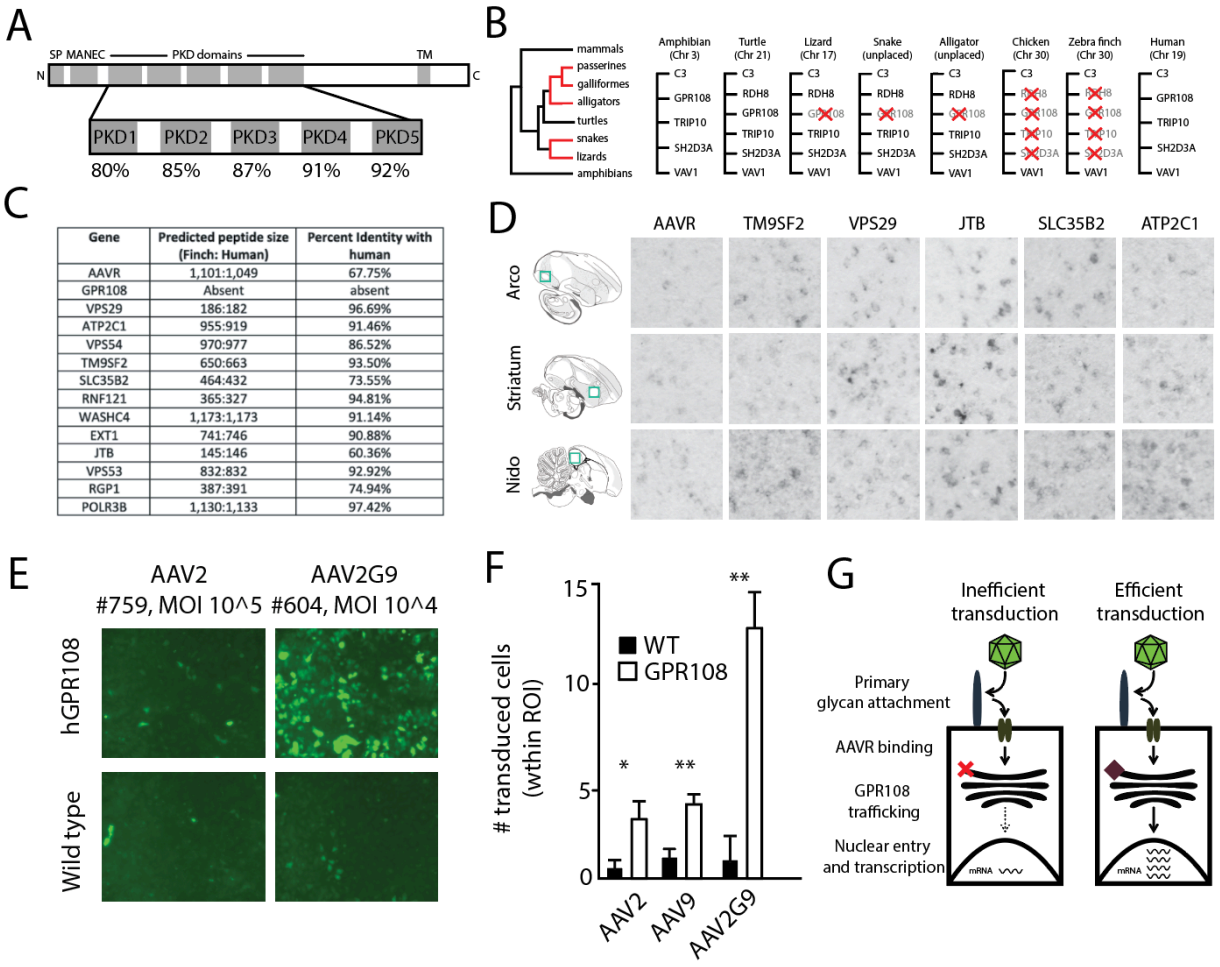
local breeder. Intraparenchymal injections of the barcoded library pdsAAV-U6-VBCx [51] (250nL) were injected into either the arcopallium or the striatum. After two weeks, finches were sacrificed via decapitation and the regions of interest were grossly dissected and flash frozen in a dry ice/isopropyl alcohol slurry.

For single virus injections, 250nl each virus was injected into the arcopallium of a female zebra finch (n=3 each). After two weeks, finches were perfused first with 0.9% NaCl followed by 3% paraformaldehyde (PFA). Brains were dissected and post-fixed in 3% PFA overnight and transferred to 30% sucrose for 2 days. Brains were then frozen in Tissue-Tek (Sakura) and sectioned at 30um on a cryostat. Immunohistochemistry was performed to enhance the GFP signal. Overnight, free floating sections were incubated in TNB (100mM Tris-HCl pH 7.4, 150mM NaCl, 0.36% w/v BSA, 1% skim milk), 0.3% Triton X-100 (Sigma), and anti-GFP 1:1000 (Abcam, ab13970). The next day, sections were washed and incubated in TNB, 0.3% Triton X-100, and Alexa Fluor 488 goat anti-chicken IgG 1:1000 (Invitrogen, A11039) for two hours. Sections were then washed, mounted on ColorFrost Plus slides (Fisher), and imaged using a Keyence BZ-X fluorescence microscope.

*AAV Barcode-Seq*

Tissue was homogenized in Trizol using a bead mill homogenizer and RNA was extracted using chloroform phase separation. AAV Barcode-Seq samples were processed as previously described (Adachi et al., 2014). Briefly, RNA was reverse transcribed using a High-Capacity cDNA Reverse Transcription Kit (Applied Biosciences). Viral barcodes (VBCs) were PCR amplified and

indexed using sample specific barcodes (SBCs). The amplicons were mixed at an equimolar ratio and sequenced on an Illumina NovaSeq 6000. Phenotypic difference was determined by quantifying the Illumina sequence read numbers for each VBC and normalizing to the library stock.




| Gene | Predicted peptide size (Finch: Human) | Percent Identity with human |
|---|---|---|
| AAVR | 1,101:1,049 | 67.75% |
| GPR108 | Absent | absent |
| VPS29 | 186:182 | 96.69% |
| ATP2C1 | 955:919 | 91.46% |
| VPS54 | 970:977 | 86.52% |
| TM9SF2 | 650:663 | 93.50% |
| SLC35B2 | 464:432 | 73.55% |
| RNF121 | 365:327 | 94.81% |
| WASHC4 | 1,173:1,173 | 91.14% |
| EXT1 | 741:746 | 90.88% |
| JTB | 145:146 | 60.36% |
| VPS53 | 832:832 | 92.92% |
| RGP1 | 387:391 | 74.94% |
| POLR3B | 1,130:1,133 | 97.42% |

Figure 5.1: Genomic loss of GPR108 in songbirds. A. Conservation of KIAA0319L (AAVR) Polycystic Kidney Disease (PKD) domains between human and zebra finch. B. Simplified cladogram (left) depicting which lineages lack GPR108 (in red) and the syntenic context of each lineage (right). C. Comparison of the predicted peptide length and amino acid conservation between human and zebra finch for the enriched genes in Pillay et al. D. Gene expression of AAV entry factors in the songbird brain. Arcopallium (Arco) is analogous to deep motor cortical layers of the mammalian cortex and Nidopallium (Nido) is analogous to superficial auditory cortical layers of the mammalian cortex. E. AAV transduction in zebra finch wild type cells and cells expressing human GPR108 (hGPR108). F. Quantification of AAV transduction in vitro in wild type and GPR108+ cells. * represents $p<0.05$, ** represents $p<0.01$, unpaired *t* test. G. Schematic representing critical stages in AAV transduction.

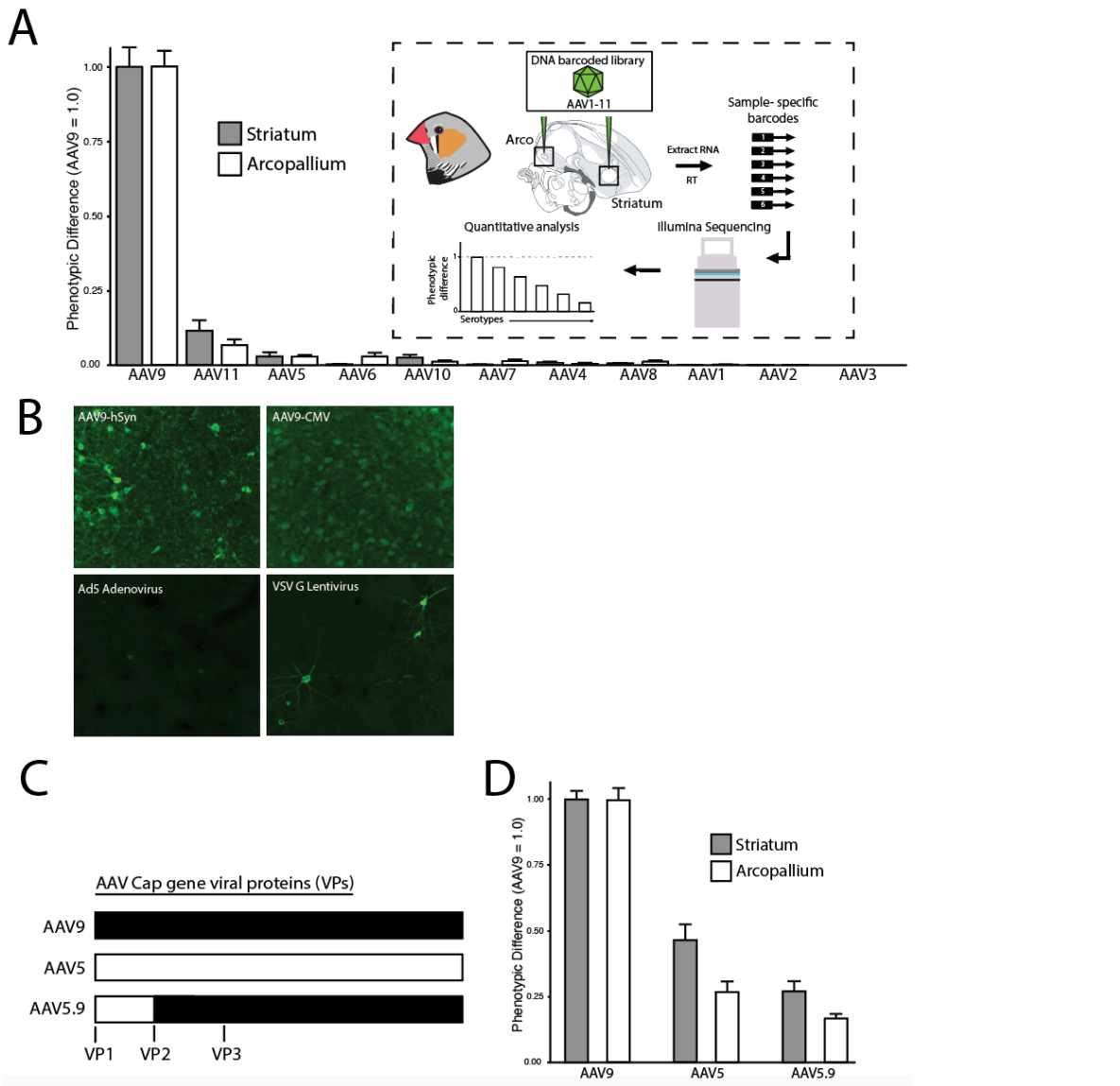



Figure 5.2: Transduction efficiency of AAVs in the songbird brain. A. Transduction efficiency of AAVs quantified using AAV Barcode-Seq. Inset shows workflow for AAV Barcode-Seq. B. Distribution of transduced cells after single vector injection in the songbird brain. C. Capsids used to test efficiency of VP1 protein in zebra finch brain transduction. D. Transduction efficiencies of capsids in C.

# Chapter 6

## Summary and Discussion

The goal of this dissertation was to describe the molecular specializations of an important brain region in the zebra finch song system. I achieved this goal through a spatial analysis of gene expression in the zebra finch arcopallium (Chapter 2) which provided context for my cell type analysis in Chapter 3. I followed this work up with an investigation into the ion channel specializations of RA which was paired a biophysical analysis of the projection neurons in RA in Chapter 4. Finally, in Chapter 5, I explored AAV transduction efficiency in the songbird brain, a surprisingly understudied topic given that there are many viral manipulation tools available to use in model organisms. These results will guide future studies to manipulate the gene expression specializations I described in the other chapters.

## 6.1 Defining molecular specializations of RA

My efforts are not the first to try to determine the how RA is unique from its surroundings. However, I believe my efforts are the most accurate description of RA gene expression specializations to date. The first gene expression specializations of RA were single gene in situ hybridizations. The songbird brain is often sectioned in the sagittal plane because a single sagittal section can contain all of the major forebrain song nuclei [152]. Despite this advantage, this orientation comes with a complication that medial and lateral information is unknown. As I

showed in Chapter 2, the coronal plane is the best for assessing specificity of RA markers so even some of the original descriptions of RA markers were misleading. In the frontal plane I discovered the close molecular relationship between RA and the adjacent AId which shows the necessity of defining and assessing context in gene expression. This theme of molecular context was critical to conceptualize past work and plan for my third chapter. Using molecular context to define gene expression specializations can be applied to any brain region in any species. The critical step is defining the most relevant region for comparison which may be straightforward or in the case of RA, be complex and require extensive analysis. Molecular context also gives clues to the evolutionary history of the brain. Just as the genome uses duplication and differentiation of genes as a mechanism of diversification, the brain may also duplicate and specialize brain pathways to execute novel behaviors [218]. The newly evolve "duplicated" brain region will carry over a large base of gene expression of the ancestral region while specializing itself with new gene expression specializations. This is what I showed with RA and AId, exhibiting similar gene expression profiles. AId of the suboscine, a closely related bird to songbirds that are not vocal learners, exhibits songbird RA/AId expression without the presence of RA. I only sampled a few of the most robust RA/AId markers so a future experiment could compare the expression profiles of suboscine and oscine AId to determine their similarity, and how they're differentiated from songbird RA. This concept of duplication and differentiation was recently shown in the evolution of cerebellar nuclei [157]. The number of cerebellar nuclei has increased throughout evolution with amphibians having 1 pair, birds and reptiles having 2 pairs, and mammals having 3 pairs of cerebellar nuclei. With each duplication comes new connectivity with the rest of the brain and differentiated cell types. The inhibitory and non-

neuronal cells were maintained whereas the excitatory neurons were differentiated [157]. This is likely a conserved mechanism for brain pathway duplications as I showed in Chapter 3 that GABAergic and non-neuronal cells were largely maintained during the evolution of RA from AId, but excitatory cells differed heavily.

Previous efforts sought to determine molecular specializations using genome-wide or near genome-wide methods [24, 25, 81]. This included a microarray, where gene-specific oligos are arrayed on a slide and RNA is hybridized to these spots [87]. A differential microarray was performed where the gene expression of RA was "subtracted" from a comparison region to attempt to define specifics. However, this comparison region selected was an auditory region [39]which produced a motor vs auditory comparison instead of a vocal-motor vs motor comparison as would be desired. When designing the experiment for Chapter 3 I had the option of repeating this type of experiment instead with a more relevant comparison region (AId) or design an experiment that determines regional expression differences but also has cell type and spatial information. I decided to perform single cell RNA seq for this experiment because I would be able to define regional differences as desired but I would be able to associate the expression differences to specific cell types (e.g. excitatory neurons, inhibitory neurons, non-neuronal cells). I combined the molecular information generated from the single cell data with a spatial analysis by in situ hybridization to define cell type and gene expression specializations in RA for the first time. Some of these specializations, like a testosterone synthesis gene, explain existing data in the field [139], but many of the defined specializations are novel with no immediate known function for how it impacts vocal production. My hope is that this dataset

inspires years of investigation into how the genes and cell types I identified enable the production of song. To this point, for future lab members and the field, I sought to develop an app to interact with this dataset. Users can query genes to determine their associated cell type.

There are still RA specializations that remain to be discovered. My single cell RNA seq analysis had relatively few cells sampled as the goal was not necessarily to discover novel cell types but instead describe the molecular phenotype of known cell types. Intuitively, the more cells that are sampled the more cell types can be resolved. Profiling many cells can discover rare cell types that drop out of the analysis of typical experiments [219]. In my analysis RA excitatory neurons were grouped as a single cell type to contrast against the excitatory neurons of the intermediate arcopallium. It is likely that there are multiple types of RA projection neurons. RA projects to both the brainstem and midbrain and it is possible that those two projections are distinct cell types [20, 74]. It is also possible that there is diversity in the brainstem projecting neurons of RA. The spatial distribution of cell types and gene expression in RA is also somewhat unexplored. Despite some reports of differential projection patterns in putative subdivisions of RA, there have been no convincing examples of within RA gene expression differences [22]. I explored this idea using bulk RNAseq and found that RA is highly homogenous in its gene expression. High cell count assays [220] and new spatial transcriptomics methods [221] are poised to help answer these questions. RA is also a highly dynamic brain region during the first three post hatch months. RA expands in size multiple times in size during acquisition of song [42]and along with that comes changes in gene expression [44]. By integrating my cell type-specific gene expression with the bulk RNAseq dataset during development I was able to infer

some changes in cell type composition of RA during this time. However, sampling RA throughout the song learning process will likely provide more information for how RA changes during development. Furthermore, RA is just one of four main forebrain nuclei for song learning and production with many more involved. The concepts I put forth in these chapters can be applied to any of the other song nuclei to define the molecular determinants of song learning and production.

## 6.2 Impacts to AAV biology

Originally, my plan was to explore a specific gene's role in the songbird brain by manipulating the expression of the gene (by expressing a gene where it is normally not found or overexpressing a dominant negative form of the gene) using viral tools. I observed that many of the commonly used viral tools in other model organisms were inefficient or underdeveloped in the songbird brain. This observation led to Chapter 5 and facilitated my interest in virology. My hope is that this chapter impacts the fields of songbird and AAV biology in multiple ways. First, through the high-throughput comparison of AAV serotypes [51] and single vector injections, I provided guidance on what viral tool to use to manipulate the brain. Second, Genome-wide screens have identified proteins critical for the transduction of AAV which are referred to as entry factors [52, 53]. Outside of generating knockout cells and animals, the impact of these genes in wild type organisms is unexplored. My finding that one of the key AAV entry factors, GPR108, is missing in songbirds and affects transduction, is the first of its kind in AAV biology. It confirms the role of GPR108 in AAV transduction but also shows that the genome of an

organism can affect viral-host interactions. Wild type AAVs have been isolated from avian species suggesting the transduction mechanisms are independent of GPR108 [213]. The example of a genomic loss is an extreme one. More likely, different tissues and cell types within tissues express AAV entry factors at varying levels. This concept of the genome/transcriptome affecting AAV transduction is largely unexplored and can have significant impacts on the development of AAV-based gene therapies.

FDA approved AAV therapies have utilized wild type serotypes thus far. The next generation of AAVs are engineered and selected for specific transduction patterns [222]. A majority of these engineered viruses have amino acids inserted into critical domains and their expression is screened in a high-throughput manner. Upon selecting enriched candidates for the specific tissue type, it is not immediately clear why the capsid produced robust transduction. In the default use case this may not be an issue but when different species, tissue types, and cell types are inaccessible, other strategies for capsid engineering may be needed. For example, in 2016 a capsid screen identified PHP.B, which contains an inserted TLAVPFK peptide at the 3-fold axis, to be able to transduce the brain 40-fold higher than its backbone serotype AAV9 following intravenous delivery [223]. This is highly desirable because the brain is difficult to access and an efficient intravenous therapy could be impactful for a variety for brain disorders. However, the robust properties of PHP.B were not observed in non-human primates or even strains of mice besides what was used in the original screen [224]. It was discovered that PHP.B was able to robustly cross the blood brain barrier by interacting with a haplotype of the Ly6a gene found in C57BL/6 mice. The selection process was successful but resulted in something truly specific to

the selection organism. The identified interaction with Ly6a allows researchers to screen the likelihood of permissive tropism in various strains of mice and species in silico [225]. This analysis is an example of the utilizing the genome to assess the transduction efficiency of a capsid.

While AAVR and GPR108 have been identified as key AAV entry factors, they are not universal. AAV4 has been shown to enter the cell independently of AAVR and AAV5 can transduce cells independently of GPR108 [53, 226]. The independence of AAV5 can actually be conferred to other capsids by swapping the VP1u domain. As discussed above, GPR108 being absent in an organism is an extreme example of the genome/transcriptome affecting transduction but this likely can impact the development of gene therapies in humans. Take for example two hypothetical cell types in the brain. Cell type 1 expresses high AAVR and high GPR108, and cell type 2 expresses high AAVR and low GPR108. Cell type 1 can likely be transduced a common serotype like AAV9 but cell type 2 cannot. Entry into cell type 2 is not the issue, it is the post entry trafficking of GPR108 that is lacking. Screening a library of capsids based on AAV9 to transduce cell type 2 likely has a ceiling of moderate transduction. Instead, taking into account of the transcriptome of the cell, a domain swapped capsid containing GPR108 independence of AAV5 with an AAV9 backbone (or even a modified AAV9) would give the highest likelihood of success. This strategy of rational design may lead to improved therapeutics. A combination of genome/transcriptome guided capsid engineering and high throughput screen capsid engineering may give the best opportunity to transduce specific populations of cells.

## 6.3 Future Directions

My dissertation work led to multiple avenues of work for immediate pursual. The first is determining how cell types, and the gene expression within each cell type, changes during vocal development. I made inferences to this question by integrating an adult cell type dataset with bulk RNA-seq from developmental timepoints [44]. My motivation for doing that analysis stemmed from the surprising finding that the developmental increase in oligodendrocytes, and presumably myelination, was not unique to male RA and was also in female RA. My analysis revealed male-specific developmental changes, primarily an increase in glial markers and a decrease in GABAergic markers. The behavioral changes that occur during vocal development (sensory phase, babbling, crystallization) have been extensively studied [9] but changes in the cell type composition of song nuclei during that time period were previously unknown. Performing single cell RNA sequencing at a higher resolution of timepoints during vocal development will directly measure how the cell type composition (e.g. proportion of each cell type) of RA changes. More timepoints sampled will capture subtle, potentially transient differences, that my analysis missed. This strategy would also determine how the gene expression profiles of individual cell types change during vocal development. An interesting timepoint to study that would require higher sampling resolution is the innervation of RA. RA receives two main inputs from song nuclei, LMAN axons which arrive pre-sensory and HVC neurons which enter RA at approximately 4 weeks post hatch [227]. The gene expression changes that occur due to HVC innervation is not known. Comparing the developing RA to AId would reveal which features of RA are developmentally distinguished from AId during vocal

development. I hypothesize that the 20dph RA and AId excitatory neurons are highly similar and that during vocal development, become differentiated.

The second follow up to my work is related to interaction with data. Interacting with single cell RNA sequencing data can be challenging because it requires some programming and molecular biology knowledge. But many neuroscientists, regardless of background, would benefit from interacting with these datasets. Interacting with datasets difficult to acquire was the motivation for the Mello lab Zebra Finch Brain Atlas (ZEBrA)[22], which is an in situ hybridization atlas akin to the Allen Institute Brain Atlas [101]. I saw the utility of my dataset for the songbird community, both for identifying novel markers and to provide context for existing gene expression profiles. To facilitate exploration, I developed an interactive application based on the R shiny platform to integrate within the ZEBrA platform. This is the first non-in situ hybridization data to be added to ZEBrA. Expanding the size of single cell datasets, brain regions represented, and degree of integration within ZEBrA will have an immediate impact on the field.

One of the most impactful parts of my dissertation was not related to identifying molecular specializations but instead the genomic loss of an AAV entry factor. As detailed in the previous chapter, I identified the viral vector that best transduces the zebra finch brain while also providing some mechanistic clues for the low transduction rate compared to mice. I showed that AAV9 is the best at transducing the zebra finch brain which can used for brain manipulation experiments. There are many suitable candidates for gene manipulation previously identified on ZEBrA and from my experiments. Beyond brain manipulation studies, my AAV work opens up avenues for basic AAV biology and host-viral interaction studies. One

outstanding question is how animals lacking GPR108 are transduced with AAV. Clearly transduction is attenuated in zebra finch cells but some cells are susceptible to transduction. Is this chance or due to a specific feature of those cells? Related to this point, there have been examples of naturally occurring AAVs isolated from species lacking GPR108, including from chicken [228]. GPR108 is critical for AAV transduction in mammals [53] but the presence of naturally occurring AAVs in GPR108-lacking species suggests that the AAV capsid evolved to differently interact in those species. Determining the residues necessary for AAV transduction, for example by a scanning alanine mutation library, could highlight species-specific viral host interactions. Similarly, determining how naturally occurring AAVs in GPR108-lacking species differ from primate-derived AAVs may give clues to how the AAV capsid can evolve to evade host factors (or lack thereof).

## 6.4 Conclusion

Songbirds have a rich behavior that allows for investigation into how the brain executes a complex sensory motor behavior. In this dissertation, I took novel approaches to address critical questions in the field of songbird neurobiology. The molecular context approach enabled the most robust description of RA specializations to date and considering the genome/transcriptome in virology has potential impacts in developing tools for basic science research and for gene therapy vector development.

# 7. References


1. Duman, D. and M. Tekin, *Autosomal recessive nonsyndromic deafness genes: a review.* Front Biosci (Landmark Ed), 2012. **17**(6): p. 2213-36.
2. Mody, M. and J.W. Belliveau, *Speech and Language Impairments in Autism: Insights from Behavior and Neuroimaging.* N Am J Med Sci (Boston), 2013. **5**(3): p. 157-161.
3. Janik, V.M. and M. Knornschild, *Vocal production learning in mammals revisited.* Philos Trans R Soc Lond B Biol Sci, 2021. **376**(1836): p. 20200244.
4. Mahrt, E.J., et al., *Engineered deafness reveals that mouse courtship vocalizations do not require auditory experience.* J Neurosci, 2013. **33**(13): p. 5573-83.
5. Hammerschmidt, K., et al., *Mice lacking the cerebral cortex develop normal song: insights into the foundations of vocal learning.* Sci Rep, 2015. **5**: p. 8808.
6. Mello, C.V., *The zebra finch, Taeniopygia guttata: an avian model for investigating the neurobiological basis of vocal learning.* Cold Spring Harb Protoc, 2014. **2014**(12): p. 1237-42.
7. Nottebohm, F. and A.P. Arnold, *Sexual dimorphism in vocal control areas of the songbird brain.* Science, 1976. **194**(4261): p. 211-3.
8. Wade, J. and A.P. Arnold, *Sexual differentiation of the zebra finch song system.* Ann N Y Acad Sci, 2004. **1016**: p. 540-59.
9. Tchernichovski, O., et al., *Dynamics of the vocal imitation process: how a zebra finch learns its song.* Science, 2001. **291**(5513): p. 2564-9.
10. Jarvis, E.D., et al., *Avian brains and a new understanding of vertebrate brain evolution.* Nat Rev Neurosci, 2005. **6**(2): p. 151-9.
11. Nottebohm, F., T.M. Stokes, and C.M. Leonard, *Central control of song in the canary, Serinus canarius.* J Comp Neurol, 1976. **165**(4): p. 457-86.
12. Karten, H.J., *Neocortical evolution: neuronal circuits arise independently of lamination.* Curr Biol, 2013. **23**(1): p. R12-5.
13. Calabrese, A. and S.M. Woolley, *Coding principles of the canonical cortical microcircuit in the avian brain.* Proc Natl Acad Sci U S A, 2015. **112**(11): p. 3517-22.
14. Stacho, M., et al., *A cortex-like canonical circuit in the avian forebrain.* Science, 2020. **369**(6511).
15. Bottjer, S.W., E.A. Miesner, and A.P. Arnold, *Forebrain lesions disrupt development but not maintenance of song in passerine birds.* Science, 1984. **224**(4651): p. 901-3.
16. Scharff, C. and F. Nottebohm, *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**(9): p. 2896-913.
17. Andalman, A.S. and M.S. Fee, *A basal ganglia-forebrain circuit in the songbird biases motor output to avoid vocal errors.* Proc Natl Acad Sci U S A, 2009. **106**(30): p. 12518-23.
18. Hahnloser, R.H., A.A. Kozhevnikov, and M.S. Fee, *An ultra-sparse code underlies the generation of neural sequences in a songbird.* Nature, 2002. **419**(6902): p. 65-70.

19. Long, M.A. and M.S. Fee, *Using temperature to analyse temporal dynamics in the songbird motor pathway.* Nature, 2008. **456**(7219): p. 189-94.
20. Vicario, D.S., *Organization of the zebra finch song control system: II. Functional organization of outputs from nucleus Robustus archistriatalis.* J Comp Neurol, 1991. **309**(4): p. 486-94.
21. Carleton, J.B., et al., *An optimized protocol for high-throughput in situ hybridization of zebra finch brain.* Cold Spring Harb Protoc, 2014. **2014**(12): p. 1249-58.
22. Lovell, P.V., et al., *ZEBrA: Zebra finch Expression Brain Atlas-A resource for comparative molecular neuroanatomy and brain evolution studies.* J Comp Neurol, 2020. **528**(12): p. 2099-2131.
23. Stark, R., M. Grzelak, and J. Hadfield, *RNA sequencing: the teenage years.* Nat Rev Genet, 2019. **20**(11): p. 631-656.
24. Pfenning, A.R., et al., *Convergent transcriptional specializations in the brains of humans and song-learning birds.* Science, 2014. **346**(6215): p. 1256846.
25. Lovell, P.V., et al., *The constitutive differential transcriptome of a brain circuit for vocal learning.* BMC Genomics, 2018. **19**(1): p. 231.
26. Macosko, E.Z., et al., *Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets.* Cell, 2015. **161**(5): p. 1202-1214.
27. Klein, A.M., et al., *Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells.* Cell, 2015. **161**(5): p. 1187-1201.
28. Cao, J., et al., *Comprehensive single-cell transcriptional profiling of a multicellular organism.* Science, 2017. **357**(6352): p. 661-667.
29. Warren, W.C., et al., *The genome of a songbird.* Nature, 2010. **464**(7289): p. 757-62.
30. Whitney, O., et al., *Core and region-enriched networks of behaviorally regulated genes and the singing genome.* Science, 2014. **346**(6215): p. 1256780.
31. Korlach, J., et al., *De novo PacBio long-read and phased avian genome assemblies correct and add to reference genes generated with intermediate and short reads.* Gigascience, 2017. **6**(10): p. 1-16.
32. Rhie, A., et al., *Towards complete and error-free genome assemblies of all vertebrate species.* Nature, 2021. **592**(7856): p. 737-746.
33. Simonyan, K., *The laryngeal motor cortex: its organization and connectivity.* Curr Opin Neurobiol, 2014. **28**: p. 15-21.
34. Mello, C.V., et al., *Molecular architecture of the zebra finch arcopallium.* J Comp Neurol, 2019. **527**(15): p. 2512-2556.
35. Feenders, G., et al., *Molecular mapping of movement-associated areas in the avian brain: a motor theory for vocal learning origin.* PLoS One, 2008. **3**(3): p. e1768.
36. Zeier, H. and H.J. Karten, *The archistriatum of the pigeon: organization of afferent and efferent connections.* Brain Res, 1971. **31**(2): p. 313-26.
37. Dugas-Ford, J., J.J. Rowell, and C.W. Ragsdale, *Cell-type homologies and the origins of the neocortex.* Proc Natl Acad Sci U S A, 2012. **109**(42): p. 16974-9.
38. Nevue, A.A., et al., *Molecular specializations of deep cortical layer analogs in songbirds.* Sci Rep, 2020. **10**(1): p. 18767.
39. Mandelblat-Cerf, Y., et al., *A role for descending auditory cortical projections in songbird vocal learning.* Elife, 2014. **3**.

40. Jarvis, E.D., *Evolution of vocal learning and spoken language.* Science, 2019. **366**(6461): p. 50-54.
41. Gahr, M., *Neural song control system of hummingbirds: comparison to swifts, vocal learning (Songbirds) and nonlearning (Suboscines) passerines, and vocal learning (Budgerigars) and nonlearning (Dove, owl, gull, quail, chicken) nonpasserines.* J Comp Neurol, 2000. **426**(2): p. 182-96.
42. Nixdorf-Bergweiler, B.E., *Divergent and parallel development in volume sizes of telencephalic song nuclei in male and female zebra finches.* J Comp Neurol, 1996. **375**(3): p. 445-56.
43. Gurney, M.E., *Hormonal control of cell form and number in the zebra finch song system.* J Neurosci, 1981. **1**(6): p. 658-73.
44. Friedrich, S.R., et al., *Emergence of sex-specific transcriptomes in a sexually dimorphic brain nucleus.* Cell Rep, 2022. **40**(5): p. 111152.
45. Zemel, B.M., et al., *Resurgent Na(+) currents promote ultrafast spiking in projection neurons that drive fine motor control.* Nat Commun, 2021. **12**(1): p. 6762.
46. Haggerty, D.L., et al., *Adeno-Associated Viral Vectors in Neuroscience Research.* Mol Ther Methods Clin Dev, 2020. **17**: p. 69-82.
47. Large, E.E., et al., *Adeno-Associated Virus (AAV) Gene Delivery: Dissecting Molecular Interactions upon Cell Entry.* Viruses, 2021. **13**(7).
48. Xiao, L., et al., *Expression of FoxP2 in the basal ganglia regulates vocal motor sequences in the adult songbird.* Nat Commun, 2021. **12**(1): p. 2617.
49. Norton, P., et al., *Differential Song Deficits after Lentivirus-Mediated Knockdown of FoxP1, FoxP2, or FoxP4 in Area X of Juvenile Zebra Finches.* J Neurosci, 2019. **39**(49): p. 9782-9796.
50. Heston, J.B. and S.A. White, *To transduce a zebra finch: interrogating behavioral mechanisms in a model system for speech.* J Comp Physiol A Neuroethol Sens Neural Behav Physiol, 2017. **203**(9): p. 691-706.
51. Adachi, K., et al., *Drawing a high-resolution functional map of adeno-associated virus capsid by massively parallel sequencing.* Nat Commun, 2014. **5**: p. 3075.
52. Pillay, S., et al., *An essential receptor for adeno-associated virus infection.* Nature, 2016. **530**(7588): p. 108-12.
53. Dudek, A.M., et al., *GPR108 Is a Highly Conserved AAV Entry Factor.* Mol Ther, 2020. **28**(2): p. 367-381.
54. Reiner, A., et al., *Revised nomenclature for avian telencephalon and some related brainstem nuclei.* J Comp Neurol, 2004. **473**(3): p. 377-414.
55. Bottjer, S.W., et al., *Axonal connections of a forebrain nucleus involved with vocal learning in zebra finches.* J Comp Neurol, 1989. **279**(2): p. 312-26.
56. Brainard, M.S. and A.J. Doupe, *What songbirds teach us about learning.* Nature, 2002. **417**(6886): p. 351-8.
57. Paton, J.A., K.R. Manogue, and F. Nottebohm, *Bilateral organization of the vocal control pathway in the budgerigar, Melopsittacus undulatus.* J Neurosci, 1981. **1**(11): p. 1279-88.
58. Durand, S.E., et al., *Vocal control pathways through the anterior forebrain of a parrot (Melopsittacus undulatus).* J Comp Neurol, 1997. **377**(2): p. 179-206.

59. Jarvis, E.D. and C.V. Mello, *Molecular mapping of brain areas involved in parrot vocal communication.* J Comp Neurol, 2000. **419**(1): p. 1-31.
60. Jarvis, E.D., et al., *Behaviourally driven gene expression reveals song nuclei in hummingbird brain.* Nature, 2000. **406**(6796): p. 628-32.
61. Doupe, A.J. and P.K. Kuhl, *Birdsong and human speech: common themes and mechanisms.* Annu Rev Neurosci, 1999. **22**: p. 567-631.
62. Wirthlin, M., et al., *A Modular Approach to Vocal Learning: Disentangling the Diversity of a Complex Behavioral Trait.* Neuron, 2019. **104**(1): p. 87-99.
63. Knornschild, M., *Vocal production learning in bats.* Curr Opin Neurobiol, 2014. **28**: p. 80-5.
64. Janik, V.M., *Cetacean vocal learning and communication.* Curr Opin Neurobiol, 2014. **28**: p. 60-5.
65. Jurgens, U., *Neuronal control of mammalian vocalization, with special reference to the squirrel monkey.* Naturwissenschaften, 1998. **85**(8): p. 376-88.
66. Hammerschmidt, K., et al., *Mice do not require auditory input for the normal development of their ultrasonic vocalizations.* BMC Neurosci, 2012. **13**: p. 40.
67. Arriaga, G., E.P. Zhou, and E.D. Jarvis, *Of mice, birds, and men: the mouse ultrasonic song system has some features similar to humans and song-learning birds.* PLoS One, 2012. **7**(10): p. e46610.
68. Kelley, D.B. and F. Nottebohm, *Projections of a telencephalic auditory nucleus-field L-in the canary.* J Comp Neurol, 1979. **183**(3): p. 455-69.
69. Margoliash, D., et al., *Distributed representation in the song system of oscines: evolutionary implications and functional consequences.* Brain Behav Evol, 1994. **44**(4-5): p. 247-64.
70. Mello, C.V., et al., *Descending auditory pathways in the adult male zebra finch (Taeniopygia guttata).* J Comp Neurol, 1998. **395**(2): p. 137-60.
71. Farries, M.A., *The avian song system in comparative perspective.* Ann N Y Acad Sci, 2004. **1016**: p. 61-76.
72. Yu, A.C. and D. Margoliash, *Temporal hierarchical control of singing in birds.* Science, 1996. **273**(5283): p. 1871-5.
73. Leonardo, A. and M.S. Fee, *Ensemble coding of vocal control in birdsong.* J Neurosci, 2005. **25**(3): p. 652-61.
74. Wild, J.M., *Descending projections of the songbird nucleus robustus archistriatalis.* J Comp Neurol, 1993. **338**(2): p. 225-41.
75. Yuan, R.C. and S.W. Bottjer, *Multidimensional Tuning in Motor Cortical Neurons during Active Behavior.* eNeuro, 2020. **7**(4).
76. Jarvis, E.D., et al., *Global view of the functional molecular organization of the avian cerebrum: mirror images and functional columns.* J Comp Neurol, 2013. **521**(16): p. 3614-65.
77. Herold, C., et al., *Transmitter receptors reveal segregation of the arcopallium/amygdala complex in pigeons (Columba livia).* J Comp Neurol, 2018. **526**(3): p. 439-466.
78. Martinez-Garcia, F., A. Martinez-Marcos, and E. Lanuza, *The pallial amygdala of amniote vertebrates: evolution of the concept, evolution of the structure.* Brain Res Bull, 2002. **57**(3-4): p. 463-9.

79. Vicario, A., et al., *Genoarchitecture of the extended amygdala in zebra finch, and expression of FoxP2 in cell corridors of different genetic profile.* Brain Struct Funct, 2017. **222**(1): p. 481-514.
80. Saldanha, C.J., et al., *Telencephalic aromatase but not a song circuit in a sub-oscine passerine, the golden collared manakin (Manacus vitellinus).* Brain Behav Evol, 2000. **56**(1): p. 29-37.
81. Hara, E., et al., *Convergent differential regulation of parvalbumin in the brains of vocal learners.* PLoS One, 2012. **7**(1): p. e29457.
82. Johnson, F., M.M. Sablan, and S.W. Bottjer, *Topographic organization of a forebrain pathway involved with vocal learning in zebra finches.* J Comp Neurol, 1995. **358**(2): p. 260-78.
83. Bottjer, S.W., J.D. Brady, and B. Cribbs, *Connections of a motor cortical region in zebra finches: relation to pathways for vocal learning.* J Comp Neurol, 2000. **420**(2): p. 244-60.
84. Jarvis, E.D. and F. Nottebohm, *Motor-driven gene expression.* Proc Natl Acad Sci U S A, 1997. **94**(8): p. 4097-102.
85. Bottjer, S.W. and B. Altenau, *Parallel pathways for vocal learning in basal ganglia of songbirds.* Nat Neurosci, 2010. **13**(2): p. 153-5.
86. Replogle, K., et al., *The Songbird Neurogenomics (SoNG) Initiative: community-based tools and strategies for study of brain gene function and evolution.* BMC Genomics, 2008. **9**: p. 131.
87. Lovell, P.V., et al., *Curation of microarray oligonucleotides and corresponding ESTs/cDNAs used for gene expression analysis in zebra finches.* BMC Res Notes, 2018. **11**(1): p. 309.
88. Konishi, M. and E. Akutagawa, *Neuronal growth, atrophy and death in a sexually dimorphic song nucleus in the zebra finch brain.* Nature, 1985. **315**(6015): p. 145-7.
89. Spiro, J.E., M.B. Dalva, and R. Mooney, *Long-range inhibition within the zebra finch song nucleus RA can coordinate the firing of multiple projection neurons.* J Neurophysiol, 1999. **81**(6): p. 3007-20.
90. Bottjer, S.W., E.A. Miesner, and A.P. Arnold, *Changes in neuronal number, density and size account for increases in volume of song-control nuclei during song development in zebra finches.* Neurosci Lett, 1986. **67**(3): p. 263-8.
91. Olveczky, B.P., et al., *Changes in the neural control of a complex motor sequence during learning.* J Neurophysiol, 2011. **106**(1): p. 386-97.
92. Tang, Y.P. and J. Wade, *Developmental changes in BDNF protein in the song control nuclei of zebra finches.* Neuroscience, 2013. **250**: p. 578-87.
93. Merullo, D.P., et al., *Neurotensin and neurotensin receptor 1 mRNA expression in song-control regions changes during development in male zebra finches.* Dev Neurobiol, 2018. **78**(7): p. 671-686.
94. Hayase, S., et al., *Vocal practice regulates singing activity-dependent genes underlying age-independent vocal learning in songbirds.* PLoS Biol, 2018. **16**(9): p. e2006537.
95. Iyengar, S., S.S. Viswanathan, and S.W. Bottjer, *Development of topography within song control circuitry of zebra finches during the sensitive period for song learning.* J Neurosci, 1999. **19**(14): p. 6037-57.

96. Aronov, D., A.S. Andalman, and M.S. Fee, *A specialized forebrain circuit for vocal babbling in the juvenile songbird.* Science, 2008. **320**(5876): p. 630-4.
97. Liu, W.C., et al., *Rudimentary substrates for vocal learning in a suboscine.* Nat Commun, 2013. **4**: p. 2082.
98. Fusani, L., et al., *Expression of androgen receptor in the brain of a sub-oscine bird with an elaborate courtship display.* Neurosci Lett, 2014. **578**: p. 61-5.
99. Wang, R., et al., *Convergent differential regulation of SLIT-ROBO axon guidance genes in the brains of vocal learners.* J Comp Neurol, 2015. **523**(6): p. 892-906.
100. Fernandez, M., et al., *Parallel organization of the avian sensorimotor arcopallium: Tectofugal visual pathway in the pigeon (Columba livia).* J Comp Neurol, 2020. **528**(4): p. 597-623.
101. Lein, E.S., et al., *Genome-wide atlas of gene expression in the adult mouse brain.* Nature, 2007. **445**(7124): p. 168-76.
102. Yuan, R.C. and S.W. Bottjer, *Differential developmental changes in cortical representations of auditory-vocal stimuli in songbirds.* J Neurophysiol, 2019. **121**(2): p. 530-548.
103. Kearney, M.G., et al., *Discrete Evaluative and Premotor Circuits Enable Vocal Learning in Songbirds.* Neuron, 2019. **104**(3): p. 559-575 e6.
104. Friedrich, S.R., et al., *Exploring the molecular basis of neuronal excitability in a vocal learner.* BMC Genomics, 2019. **20**(1): p. 629.
105. Clayton, D.F., et al., *Conservation and expression of IQ-domain-containing calpacitin gene products (neuromodulin/GAP-43, neurogranin/RC3) in the adult and developing oscine song control system.* Dev Neurobiol, 2009. **69**(2-3): p. 124-40.
106. George, J.M., et al., *Characterization of a novel protein regulated during the critical period for song learning in the zebra finch.* Neuron, 1995. **15**(2): p. 361-72.
107. Spillantini, M.G., et al., *Alpha-synuclein in Lewy bodies.* Nature, 1997. **388**(6645): p. 839-40.
108. Singleton, A.B., et al., *alpha-Synuclein locus triplication causes Parkinson's disease.* Science, 2003. **302**(5646): p. 841.
109. Masliah, E., et al., *beta-amyloid peptides enhance alpha-synuclein accumulation and neuronal deficits in a transgenic mouse model linking Alzheimer's disease and Parkinson's disease.* Proc Natl Acad Sci U S A, 2001. **98**(21): p. 12245-50.
110. Scherzer, C.R., et al., *GATA transcription factors directly regulate the Parkinson's disease-linked gene alpha-synuclein.* Proc Natl Acad Sci U S A, 2008. **105**(31): p. 10907-12.
111. Lei, H., et al., *Axon guidance pathways served as common targets for human speech/language evolution and related disorders.* Brain Lang, 2017. **174**: p. 1-8.
112. Petkov, C.I. and E.D. Jarvis, *Birds, primates, and spoken language origins: behavioral phenotypes and neurobiological substrates.* Front Evol Neurosci, 2012. **4**: p. 12.
113. Sin, C., H. Li, and D.A. Crawford, *Transcriptional regulation by FOXP1, FOXP2, and FOXP4 dimerization.* J Mol Neurosci, 2015. **55**(2): p. 437-48.
114. Wirthlin, M., et al., *Parrot Genomes and the Evolution of Heightened Longevity and Cognition.* Curr Biol, 2018. **28**(24): p. 4001-4008 e7.

115. Winberg, M.L., et al., *Plexin A is a neuronal semaphorin receptor that controls axon guidance.* Cell, 1998. **95**(7): p. 903-16.
116. Mi, S., et al., *LINGO-1 is a component of the Nogo-66 receptor/p75 signaling complex.* Nat Neurosci, 2004. **7**(3): p. 221-8.
117. Fournier, A.E., T. GrandPre, and S.M. Strittmatter, *Identification of a receptor mediating Nogo-66 inhibition of axonal regeneration.* Nature, 2001. **409**(6818): p. 341-6.
118. Lovell, P.V., J.B. Carleton, and C.V. Mello, *Genomics analysis of potassium channel genes in songbirds reveals molecular specializations of brain circuits for the maintenance and production of learned vocalizations.* BMC Genomics, 2013. **14**: p. 470.
119. Ross, M.T., et al., *Neuronal Intrinsic Physiology Changes During Development of a Learned Behavior.* eNeuro, 2017. **4**(5).
120. Daou, A. and D. Margoliash, *Intrinsic neuronal properties represent song and error in zebra finch vocal learning.* Nat Commun, 2020. **11**(1): p. 952.
121. Delvendahl, I., et al., *Plasticity of motor threshold and motor-evoked potential amplitude--a model of intrinsic and synaptic plasticity in human motor cortex?* Brain Stimul, 2012. **5**(4): p. 586-93.
122. Viiri, K.M., et al., *DNA-binding and -bending activities of SAP30L and SAP30 are mediated by a zinc-dependent module and monophosphoinositides.* Mol Cell Biol, 2009. **29**(2): p. 342-56.
123. Xie, T., et al., *Structure of the 30-kDa Sin3-associated protein (SAP30) in complex with the mammalian Sin3A corepressor and its role in nucleic acid binding.* J Biol Chem, 2011. **286**(31): p. 27814-24.
124. Kang, H.J., et al., *Spatio-temporal transcriptome of the human brain.* Nature, 2011. **478**(7370): p. 483-9.
125. Bormuth, I., et al., *Neuronal basic helix-loop-helix proteins Neurod2/6 regulate cortical commissure formation before midline interactions.* J Neurosci, 2013. **33**(2): p. 641-51.
126. Nakagawa, Y. and D.D. O'Leary, *Dynamic patterned expression of orphan nuclear receptor genes RORalpha and RORbeta in developing mouse forebrain.* Dev Neurosci, 2003. **25**(2-4): p. 234-44.
127. Sarachana, T. and V.W. Hu, *Genome-wide identification of transcriptional targets of RORA reveals direct regulation of multiple genes associated with autism spectrum disorder.* Mol Autism, 2013. **4**(1): p. 14.
128. Vates, G.E., et al., *Auditory pathways of caudal telencephalon and their relation to the song system of adult male zebra finches.* J Comp Neurol, 1996. **366**(4): p. 613-42.
129. Cheng, M., et al., *Nucleus taenia of the amygdala of birds: anatomical and functional studies in ring doves (Streptopelia risoria) and European starlings (Sturnus vulgaris).* Brain Behav Evol, 1999. **53**(5-6): p. 243-70.
130. Arendt, D., et al., *The origin and evolution of cell types.* Nat Rev Genet, 2016. **17**(12): p. 744-757.
131. Arendt, D., *The evolution of cell types in animals: emerging principles from molecular studies.* Nat Rev Genet, 2008. **9**(11): p. 868-82.
132. Zeier, H.J. and H.J. Karten, *Connections of the anterior commissure in the pigeon (Columba livia).* J Comp Neurol, 1973. **150**(2): p. 201-16.

133. Colquitt, B.M., et al., *Cellular transcriptomics reveals evolutionary identities of songbird vocal circuits.* Science, 2021. **371**(6530).
134. Zheng, G.X., et al., *Massively parallel digital transcriptional profiling of single cells.* Nat Commun, 2017. **8**: p. 14049.
135. Zhang, Y., et al., *An RNA-sequencing transcriptome and splicing database of glia, neurons, and vascular cells of the cerebral cortex.* J Neurosci, 2014. **34**(36): p. 11929-47.
136. Tasic, B., et al., *Adult mouse cortical cell taxonomy revealed by single cell transcriptomics.* Nat Neurosci, 2016. **19**(2): p. 335-46.
137. von Bartheld, C.S., J. Bahney, and S. Herculano-Houzel, *The search for true numbers of neurons and glial cells in the human brain: A review of 150 years of cell counting.* J Comp Neurol, 2016. **524**(18): p. 3865-3895.
138. Gurney, M.E. and M. Konishi, *Hormone-induced sexual differentiation of brain and behavior in zebra finches.* Science, 1980. **208**(4450): p. 1380-3.
139. Grisham, W., et al., *A putative 5 alpha-reductase inhibitor demasculinizes portions of the zebra finch song system.* Brain Res, 1997. **750**(1-2): p. 122-8.
140. Song, S.J., et al., *LGR5/GPR49 is implicated in motor neuron specification in nervous system.* Neurosci Lett, 2015. **584**: p. 135-40.
141. Carmon, K.S., et al., *R-spondins function as ligands of the orphan receptors LGR4 and LGR5 to regulate Wnt/beta-catenin signaling.* Proc Natl Acad Sci U S A, 2011. **108**(28): p. 11452-7.
142. McLeod, F. and P.C. Salinas, *Wnt proteins as modulators of synaptic plasticity.* Curr Opin Neurobiol, 2018. **53**: p. 90-95.
143. Olson, C.R., L.K. Hodges, and C.V. Mello, *Dynamic gene expression in the song system of zebra finches during the song learning period.* Dev Neurobiol, 2015. **75**(12): p. 1315-38.
144. Holzenberger, M., et al., *Selective expression of insulin-like growth factor II in the songbird brain.* J Neurosci, 1997. **17**(18): p. 6974-87.
145. Roberts, T.F., et al., *Telencephalic neurons monosynaptically link brainstem and forebrain premotor networks necessary for song.* J Neurosci, 2008. **28**(13): p. 3479-89.
146. Bakken, T.E., et al., *Comparative cellular analysis of motor cortex in human, marmoset and mouse.* Nature, 2021. **598**(7879): p. 111-119.
147. Balmer, T.S., et al., *Modulation of perineuronal nets and parvalbumin with developmental song learning.* J Neurosci, 2009. **29**(41): p. 12878-85.
148. Wild, J.M., M.N. Williams, and R.A. Suthers, *Parvalbumin-positive projection neurons characterise the vocal premotor pathway in male, but not female, zebra finches.* Brain Res, 2001. **917**(2): p. 235-52.
149. Pinaud, R. and C.V. Mello, *GABA immunoreactivity in auditory and song control brain areas of zebra finches.* J Chem Neuroanat, 2007. **34**(1-2): p. 1-21.
150. Dimidschstein, J., et al., *A viral strategy for targeting and manipulating interneurons across vertebrate species.* Nat Neurosci, 2016. **19**(12): p. 1743-1749.
151. Champoux, K.L., K.E. Miller, and D.J. Perkel, *Differential development of myelin in zebra finch song nuclei.* J Comp Neurol, 2021. **529**(6): p. 1255-1265.
152. Karten, H.J., et al., *Digital atlas of the zebra finch (Taeniopygia guttata) brain: a high-resolution photo atlas.* J Comp Neurol, 2013. **521**(16): p. 3702-15.

153. Adret, P. and D. Margoliash, *Metabolic and neural activity in the song system nucleus robustus archistriatalis: effect of age and gender.* J Comp Neurol, 2002. **454**(4): p. 409-23.
154. Safra, N., et al., *Genome-wide association mapping in dogs enables identification of the homeobox gene, NKX2-8, as a genetic component of neural tube defects in humans.* PLoS Genet, 2013. **9**(7): p. e1003646.
155. Lodato, S. and P. Arlotta, *Generating neuronal diversity in the mammalian cerebral cortex.* Annu Rev Cell Dev Biol, 2015. **31**: p. 699-720.
156. Santello, M., N. Toni, and A. Volterra, *Astrocyte function from information processing to cognition and cognitive impairment.* Nat Neurosci, 2019. **22**(2): p. 154-166.
157. Kebschull, J.M., et al., *Cerebellar nuclei evolved by repeatedly duplicating a conserved cell-type set.* Science, 2020. **370**(6523).
158. Stuart, T., et al., *Comprehensive Integration of Single-Cell Data.* Cell, 2019. **177**(7): p. 1888-1902 e21.
159. Susaki, E.A., et al., *Advanced CUBIC protocols for whole-brain and whole-body clearing and imaging.* Nat Protoc, 2015. **10**(11): p. 1709-27.
160. Lee, E., et al., *ACT-PRESTO: Rapid and consistent tissue clearing and labeling method for 3-dimensional (3D) imaging.* Sci Rep, 2016. **6**: p. 18631.
161. Ebbesen, C.L. and M. Brecht, *Motor cortex - to act or not to act?* Nat Rev Neurosci, 2017. **18**(11): p. 694-705.
162. Suter, B.A., M. Migliore, and G.M. Shepherd, *Intrinsic electrophysiology of mouse corticospinal neurons: a class-specific triad of spike-related properties.* Cereb Cortex, 2013. **23**(8): p. 1965-77.
163. Oswald, M.J., et al., *Diversity of layer 5 projection neurons in the mouse motor cortex.* Front Cell Neurosci, 2013. **7**: p. 174.
164. Paz, J.T., et al., *Multiple forms of activity-dependent intrinsic plasticity in layer V cortical neurones in vivo.* J Physiol, 2009. **587**(Pt 13): p. 3189-205.
165. Shim, H.G., Y.S. Lee, and S.J. Kim, *The Emerging Concept of Intrinsic Plasticity: Activity-dependent Modulation of Intrinsic Excitability in Cerebellar Purkinje Cells and Motor Learning.* Exp Neurobiol, 2018. **27**(3): p. 139-154.
166. Eshra, A., P. Hirrlinger, and S. Hallermann, *Enriched Environment Shortens the Duration of Action Potentials in Cerebellar Granule Cells.* Front Cell Neurosci, 2019. **13**: p. 289.
167. Nakajima, K., et al., *Striking differences in transmission of corticospinal excitation to upper limb motoneurons in two primate species.* J Neurophysiol, 2000. **84**(2): p. 698-709.
168. Vigneswaran, G., A. Kraskov, and R.N. Lemon, *Large identified pyramidal cells in macaque motor and premotor cortex exhibit "thin spikes": implications for cell type classification.* J Neurosci, 2011. **31**(40): p. 14235-42.
169. Soares, D., et al., *Expression of Kv3.1b potassium channel is widespread in macaque motor cortex pyramidal cells: A histological comparison between rat and macaque.* J Comp Neurol, 2017. **525**(9): p. 2164-2174.
170. Gunturkun, O. and T. Bugnyar, *Cognition without Cortex.* Trends Cogn Sci, 2016. **20**(4): p. 291-303.
171. Clayton, N.S. and N.J. Emery, *Avian Models for Human Cognitive Neuroscience: A Proposal.* Neuron, 2015. **86**(6): p. 1330-42.

172. Karten, H.J., *Vertebrate brains and evolutionary connectomics: on the origins of the mammalian 'neocortex'.* Philos Trans R Soc Lond B Biol Sci, 2015. **370**(1684).
173. Schmidt, M.F. and J. Martin Wild, *The respiratory-vocal system of songbirds: anatomy, physiology, and neural control.* Prog Brain Res, 2014. **212**: p. 297-335.
174. Mao, B.Q., et al., *Dynamics of spontaneous activity in neocortical slices.* Neuron, 2001. **32**(5): p. 883-98.
175. Zhang, Z.W., *Maturation of layer V pyramidal neurons in the rat prefrontal cortex: intrinsic properties and synaptic function.* J Neurophysiol, 2004. **91**(3): p. 1171-82.
176. Beaulieu-Laroche, L., et al., *Enhanced Dendritic Compartmentalization in Human Cortical Neurons.* Cell, 2018. **175**(3): p. 643-651 e14.
177. Elemans, C.P., et al., *Superfast vocal muscles control song production in songbirds.* PLoS One, 2008. **3**(7): p. e2581.
178. Goller, F. and R.A. Suthers, *Role of syringeal muscles in gating airflow and sound production in singing brown thrashers.* J Neurophysiol, 1996. **75**(2): p. 867-76.
179. Vicario, D.S., *Contributions of syringeal muscles to respiration and vocalization in the zebra finch.* J Neurobiol, 1991. **22**(1): p. 63-73.
180. Chi, Z. and D. Margoliash, *Temporal precision and temporal drift in brain and behavior of zebra finch song.* Neuron, 2001. **32**(5): p. 899-910.
181. Adam, I. and C.P.H. Elemans, *Increasing Muscle Speed Drives Changes in the Neuromuscular Transform of Motor Commands during Postnatal Development in Songbirds.* J Neurosci, 2020. **40**(35): p. 6722-6731.
182. Gallant, J.R. and L.A. O'Connell, *Studying convergent evolution to relate genotype to behavioral phenotype.* J Exp Biol, 2020. **223**(Pt Suppl 1).
183. Zakon, H.H., et al., *Sodium channel genes and the evolution of diversity in communication signals of electric fishes: convergent molecular evolution.* Proc Natl Acad Sci U S A, 2006. **103**(10): p. 3675-80.
184. McAnelly, M.L. and H.H. Zakon, *Coregulation of voltage-dependent kinetics of Na(+) and K(+) currents in electric organ.* J Neurosci, 2000. **20**(9): p. 3408-14.
185. Markham, M.R., L.K. Kaczmarek, and H.H. Zakon, *A sodium-activated potassium channel supports high-frequency firing and reduces energetic costs during rapid modulations of action potential amplitude.* J Neurophysiol, 2013. **109**(7): p. 1713-23.
186. Akemann, W. and T. Knopfel, *Interaction of Kv3 potassium channels and resurgent sodium current influences the rate of spontaneous firing of Purkinje neurons.* J Neurosci, 2006. **26**(17): p. 4602-12.
187. Raman, I.M. and B.P. Bean, *Resurgent sodium current and action potential formation in dissociated cerebellar Purkinje neurons.* J Neurosci, 1997. **17**(12): p. 4517-26.
188. Grieco, T.M., et al., *Open-channel block by the cytoplasmic tail of sodium channel beta4 as a mechanism for resurgent sodium current.* Neuron, 2005. **45**(2): p. 233-44.
189. Lewis, A.H. and I.M. Raman, *Resurgent current of voltage-gated Na(+) channels.* J Physiol, 2014. **592**(22): p. 4825-38.
190. Hong, H., et al., *Resurgent sodium current promotes action potential firing in the avian auditory brainstem.* J Physiol, 2018. **596**(3): p. 423-443.

191. Gobes, S.M.H., R.B. Jennings, and R.K. Maeda, *The sensitive period for auditory-vocal learning in the zebra finch: Consequences of limited-model availability and multiple-tutor paradigms on song imitation.* Behav Processes, 2019. **163**: p. 5-12.
192. Garst-Orozco, J., B. Babadi, and B.P. Olveczky, *A neural circuit mechanism for regulating vocal variability during song learning in zebra finches.* Elife, 2014. **3**: p. e03697.
193. Kittelberger, J.M. and R. Mooney, *Lesions of an avian forebrain nucleus that disrupt song development alter synaptic connectivity and transmission in the vocal premotor pathway.* J Neurosci, 1999. **19**(21): p. 9385-98.
194. Stark, L.L. and D.J. Perkel, *Two-stage, input-specific synaptic maturation in a nucleus essential for vocal production in the zebra finch.* J Neurosci, 1999. **19**(20): p. 9107-16.
195. White, S.A., F.S. Livingston, and R. Mooney, *Androgens modulate NMDA receptor-mediated EPSCs in the zebra finch song system.* J Neurophysiol, 1999. **82**(5): p. 2221-34.
196. Kaczmarek, L.K. and Y. Zhang, *Kv3 Channels: Enablers of Rapid Firing, Neurotransmitter Release, and Neuronal Endurance.* Physiol Rev, 2017. **97**(4): p. 1431-1468.
197. Rudy, B. and C.J. McBain, *Kv3 channels: voltage-gated K+ channels designed for high-frequency repetitive firing.* Trends Neurosci, 2001. **24**(9): p. 517-26.
198. Parameshwaran, S., C.E. Carr, and T.M. Perney, *Expression of the Kv3.1 potassium channel in the avian auditory brainstem.* J Neurosci, 2001. **21**(2): p. 485-94.
199. Li, W., L.K. Kaczmarek, and T.M. Perney, *Localization of two high-threshold potassium channel subunits in the rat central auditory system.* J Comp Neurol, 2001. **437**(2): p. 196-218.
200. Carr, C.E. and D. Soares, *Evolutionary convergence and shared computational principles in the auditory system.* Brain Behav Evol, 2002. **59**(5-6): p. 294-311.
201. Spool, J.A., et al., *Genetically identified neurons in avian auditory pallium mirror core principles of their mammalian counterparts.* Curr Biol, 2021. **31**(13): p. 2831-2843 e6.
202. Chen, W., et al., *Electrophysiological and morphological properties of pyramidal and nonpyramidal neurons in the cat motor cortex in vitro.* Neuroscience, 1996. **73**(1): p. 39-55.
203. Lacey, M.G., et al., *Spike firing and IPSPs in layer V pyramidal neurons during beta oscillations in rat primary motor cortex (M1) in vitro.* PLoS One, 2014. **9**(1): p. e85109.
204. Ichinohe, N., et al., *A voltage-gated potassium channel, Kv3.1b, is expressed by a subpopulation of large pyramidal neurons in layer 5 of the macaque monkey cortex.* Neuroscience, 2004. **129**(1): p. 179-85.
205. Network, B.I.C.C., *A multimodal cell census and atlas of the mammalian primary motor cortex.* Nature, 2021. **598**(7879): p. 86-102.
206. Rivara, C.B., et al., *Stereologic characterization and spatial distribution patterns of Betz cells in the human primary motor cortex.* Anat Rec A Discov Mol Cell Evol Biol, 2003. **270**(2): p. 137-51.
207. Lassek, A.M., *The human pyramidal tract: II. A numberical investigation of Betz cells of the motor area.* Archives of Neurology & Psychiatry, 1940. **44**(4): p. 7.
208. Tomasevic, L., et al., *Relationship between high-frequency activity in the cortical sensory and the motor hand areas, and their myelin content.* Brain Stimul, 2022. **15**(3): p. 717-726.

209. Replogle, K., et al., *The Songbird Neurogenomics (SoNG) Initiative: community-based tools and strategies for study of brain gene function and evolution.* BMC Genomics, 2008. **9**(1): p. 131.
210. Schindelin, J., et al., *Fiji: an open-source platform for biological-image analysis.* Nat Methods, 2012. **9**(7): p. 676-82.
211. Meisen, W.H., et al., *Pooled Screens Identify GPR108 and TM9SF2 as Host Cell Factors Critical for AAV Transduction.* Mol Ther Methods Clin Dev, 2020. **17**: p. 601-611.
212. During, D.N., et al., *Fast Retrograde Access to Projection Neuron Circuits Underlying Vocal Learning in Songbirds.* Cell Rep, 2020. **33**(6): p. 108364.
213. Matsui, R., Y. Tanabe, and D. Watanabe, *Avian adeno-associated virus vector efficiently transduces neurons in the embryonic and post-embryonic chicken brain.* PLoS One, 2012. **7**(11): p. e48730.
214. Biegler, M.T., et al., *Induction of an immortalized songbird cell line allows for gene characterization and knockout by CRISPR-Cas9.* Sci Rep, 2022. **12**(1): p. 4369.
215. Velho, T.A.F., et al., *Divergent low-density lipoprotein receptor (LDLR) linked to low VSV G-dependent viral infectivity and unique serum lipid profile in zebra finches.* Proc Natl Acad Sci U S A, 2021. **118**(18).
216. Kowarz, E., D. Loscher, and R. Marschalek, *Optimized Sleeping Beauty transposons rapidly generate stable transgenic cell lines.* Biotechnol J, 2015. **10**(4): p. 647-53.
217. Mates, L., et al., *Molecular evolution of a novel hyperactive Sleeping Beauty transposase enables robust stable gene transfer in vertebrates.* Nat Genet, 2009. **41**(6): p. 753-61.
218. Chakraborty, M. and E.D. Jarvis, *Brain evolution by brain pathway duplication.* Philos Trans R Soc Lond B Biol Sci, 2015. **370**(1684).
219. Saunders, A., et al., *Molecular Diversity and Specializations among the Cells of the Adult Mouse Brain.* Cell, 2018. **174**(4): p. 1015-1030 e16.
220. Datlinger, P., et al., *Ultra-high-throughput single-cell RNA sequencing and perturbation screening with combinatorial fluidic indexing.* Nat Methods, 2021. **18**(6): p. 635-642.
221. Stickels, R.R., et al., *Highly sensitive spatial transcriptomics at near-cellular resolution with Slide-seqV2.* Nat Biotechnol, 2021. **39**(3): p. 313-319.
222. Challis, R.C., et al., *Systemic AAV vectors for widespread and targeted gene delivery in rodents.* Nat Protoc, 2019. **14**(2): p. 379-414.
223. Deverman, B.E., et al., *Cre-dependent selection yields AAV variants for widespread gene transfer to the adult brain.* Nat Biotechnol, 2016. **34**(2): p. 204-9.
224. Hordeaux, J., et al., *The Neurotropic Properties of AAV-PHP.B Are Limited to C57BL/6J Mice.* Mol Ther, 2018. **26**(3): p. 664-668.
225. Huang, Q., et al., *Delivering genes across the blood-brain barrier: LY6A, a novel cellular receptor for AAV-PHP.B capsids.* PLoS One, 2019. **14**(11): p. e0225206.
226. Dudek, A.M., et al., *An Alternate Route for Adeno-associated Virus (AAV) Entry Independent of AAV Receptor.* J Virol, 2018. **92**(7).
227. Mooney, R. and M. Rao, *Waiting periods versus early innervation: the development of axonal connections in the zebra finch song system.* J Neurosci, 1994. **14**(11 Pt 1): p. 6532-43.
228. Bossis, I. and J.A. Chiorini, *Cloning of an avian adeno-associated virus (AAAV) and generation of recombinant AAAV particles.* J Virol, 2003. **77**(12): p. 6799-810.