A sex chromosome drives the emergence of vocal learning following hormonal manipulation in female zebra finches

Matthew Davenport[1] & Ha Na Choe[2], Hiroaki Matsunami[2, 3], Erich D. Jarvis[1,4]

1: Laboratory of Language Neurogenetics, The Rockefeller University, New York, NY, USA.
2: Department of Molecular Genetics and Microbiology,
3: Department of Neurobiology, Duke Institute for Brain Sciences, Duke University School of Medicine, Durham, NC, USA
4: Howard Hughes Medical Institute, Chevy Chase, MD

Zebra finches are sexually dimorphic vocal learners. Males learn to sing by imitating mature conspecifics, but females do not. The lack of vocal learning in females is associated with anatomical differences in the neural circuits responsible for vocal learning, including the atrophy of several brain regions during development[1]. However, this atrophy can be prevented and song learning retained in females after pharmacological estrogen treatment[2–4]. Little is known about the genetic machinery controlling this sex and estrogen responsive song system development. To screen for drivers, we performed an unbiased analysis of transcriptomes from song control nuclei and surrounding motor regions in zebra finches of either sex treated with 17-β-estradiol or vehicle until sacrifice on day 30, when divergence between the sexes is anatomically apparent. Utilizing the newly assembled autosomes and sex chromosomes from the zebra finch Vertebrate Genomes Project assemblies[5], we identified correlated gene modules that were associated to song nuclei in a sex and estradiol dependent manner. Female estradiol treated HVC, in the vocal learning circuit, acquired the smallest of the modular specializations observed in male HVC. This module was enriched for genes governing anatomical development, and it's specilization was dispraportionately influenced by the expression of Z sex chromosome transcripts in HVC. We propose that vocal learning may be prevented in female zebra finches via the suppression of an estrogen inducible Z chromosome *cis*-acting regulatory element.

Main

Vocal learning is the ability to imitate sounds. In humans, vocal learning is a necessary and specialized component for spoken language and song. Though no animal produces human-like language, the most similar phenotypes are found in 7 non-human clades of vocal learners, four mammalian and three avian, each having independently evolved the trait[6]. Oscine songbirds have proved to be the most tractable for studying vocal learning in the lab, with much of the field focusing on the Australian zebra finch (*Taeniopygia guttata*). Despite ~300 million years of separation since a common ancestor[7,8], there is remarkable evolutionary convergence between songbird and human vocal learning in terms of the behavioral progression, the developmental effects of deafening, the anatomical connectivity of vocal-motor learning circuits, the sites of accelerated evolution within the genome, and even the specific genes that mark song and speech circuits with higher or lower specialized expression against the surrounding motor control circuits[6,9–15]. Unlike in humans, however, vocal learning is strongly sexually dimorphic in zebra finches[1]. Male zebra finches learn to produce a species appropriate song by imitating mature conspecifics during juvenile development, while females are limited to producing innate calls[16].

The adult male zebra finch brain song system includes four major interconnected telencephalic song control nuclei; HVC (proper name) in the dorsal nidopallium (DN); the lateral magnocellular nucleus of the anterior nidopallium (LMAN) found in the anterior nidopallium (AN); the robust nucleus of the arcopallium (RA) in the lateral intermediate arcopallium (LAI); and Area X in the striatum (Str; **Fig. 1a**)[11]. During juvenile development in females HVC and RA atrophy, the striatal song nucleus Area X never appears, and HVC fails to form synapses in RA[1,17–24]. Amazingly, female zebra finches treated with estrogen or a synthetic analog at an early age do not exhibit song system atrophy and instead form a functional neural circuit with all the anatomical components and connections seen in males[2–4,25]. This "masculinized" song system allows estrogen supplemented females to vocally imitate adult male conspecifics, though not with the same accuracy as males[2,4,26]. Interestingly, lesioning female HVC prevents estrogen dependent anatomical "masculinization" of its postsynaptic targets, RA and Area X[27].

The genetic basis of estrogen sensitive, sexually dimorphic zebra finch vocal learning remains largely unknown. However, the examination of a rare gynandromorphic zebra finch with lateralized sex chromosome composition indicates that genetically male Area X and HVC are larger than their female analogs independent of gonadal hormone production, implicating sex chromosome gene expression within the song system[28]. Unlike the mammalian X and Y, birds use the Z and W sex determination system where females are hemizygotic (ZW), and males homozygotic (ZZ)[29]. The relevant transcriptional machinery must also be set up by posthatch day 30 (PHD30), after which estrogen fails to masculinize female brains or behavior and the male song system enlarges while the female song system atrophies[2,17,19,30]. To screen for genetic drivers and locate their action within the song system, we performed an unbiased analysis of transcriptomes from song system components and surrounding motor control regions in zebra finches of either sex treated with 17-β-estradiol (E2) or vehicle until sacrifice at PHD30. We used a new zebra finch genome assembly and annotation produced by the Vertebrate Genomes Project containing both the Z and W[5], to identify candidate drivers of vocal learning loss in females from the sex chromosomes.

Identification of gene modules

We used RNA-Seq data from a previous study on the effects of E2 manipulation on the song system which was analyzed using a differential expression approach after mapping reads to an older genome lacking the W sex chromosome[26]. We first re-mapped RNAseq reads to the new assembly from the four major song nuclei (HVC, LMAN, RA, and Area X) and their adjacent non-vocal motor surrounds (DN, AN, LAI, and Str respectively; **Fig. 1a**), taken from quiet males and females at 30 days old; treated with either E2 or a vehicle control since hatch (**Fig. 1b**)[26]. The non-vocal surrounds allow us to examine specialized up- or down-regulated gene expression patterns in the song control nuclei[12,13].

Within the mapped gene expression data, we asked the following five questions: “What were the specialized gene expression profiles of males at PHD30?”; “Which female song nuclei were specialized by E2?”; “What were the gene networks implicated in that specialization?”; “Which genes most strongly drove this specialization?”; and “Are the sex chromosomes involved, and if so how?”. To do this in an unbiased way, we performed weighted gene correlation network analysis (WGCNA) to first decompose the transcriptome into actively expressed gene modules by performing iterative hierarchical clustering on the matrix of gene-to-gene correlations in expression across samples[31]. Of the ~21,000 annotated zebra finch genes, 13,220 were expressed in the finch telencephalon and assigned to one of 14 co-expression modules, a comparable number of genes to what we have seen expressed in adult zebra finch telencephalon[14,32]. Two of the modules clearly marked single samples (**Fig. S2**), indicative of technical overfitting, and were excluded from further analysis. The remaining 12 modules, including 12,444 genes, were dynamically expressed across brain regions and treatment groups, which we lettered in descending order of size A through L, containing from 4,890 to 127 constituent genes (**Fig. 1c**).

PHD30 female HVC and Area X lack specialized gene modules

To answer our first question of modules specialized to male song nuclei, we first calculated the module eigengene (MEG) for each of our modules (**Fig. 1d**). MEGs are the first principle component of variance of all genes in a module and can be thought of as an aggregate measure of expression for each module’s expression across samples. We then associated gene modules to song nuclei by correlating MEG expression song system membership for each nucleus against its respective surround (e.g. RA vs. LAI) in vocal learning males (**Fig. 2a**). As expected, the male song system was strongly specialized in each of the four nuclei. Male HVC was specialized relative to DN by the significant differential expression of modules B,C,F, and G; LMAN was specialized relative to the surrounding AN by modules B,F,G, and I; RA was specialized relative to LAI by modules A,C, and G; and Area X was specialized relative to the Str by modules C,E,F,G, and I. To address our second question, we found that in vehicle-treated females, LMAN and RA exhibited similar gene module specialization patterns observed in males, appearing well set up for vocal learning even in the absence of E2. In contrast, neither the female vehicle HVC nor Area X showed significant gene module specializations against their surrounds, making them more likely sites of E2 action (**Fig. 2b**).

E2 sensitive HVC module for tissue expansion

To answer our third question and determine whether exogenous E2 caused specialized gene module expression within the song system in females, we compared MEG expressions in female-E2-HVC samples to the combination of female-E2-DN, female-Veh-DN, and female-Veh-HVC samples. We identified a single module whose expression was specialized in HVC of E2-treated females, and it was the same module G found in male HVC (**Fig. 2c**). This was the smallest of the four modules specialized in male HVC relative to DN, containing 344 genes (**Table S1**). To understand the biological function of this gene module, we mapped the zebra finch genes to their 1:1 human orthologs where possible and then used human gene annotation to examine the Gene Ontology (GO) functions enriched within the module G constituent genes (**Table 1**). We found significantly enriched GO terms, which included “DNA binding transcription factor activity”, “cell differentiation”, “anatomical morphogenesis”, “cell to cell signaling”, and “positive regulation of multicellular organism growth”, indicating that module G genes likely enact a proliferative developmental program. Other significantly enriched terms such as “extracellular matrix structural component”, “external side of the plasma membrane”, and “extracellular space” indicate that this module may also act to restructure the extracellular matrix, perhaps to accommodate new cells.

We also compared female-E2-Area X samples to all other female samples from the striatum, and found specialized expression module G, as well as modules C and F (**Fig. 2c**). The strong signal here is consistent with female Area X being completely absent without E2 and its presence depending on HVC masculinization by E2[27]. For RA of E2 treated females we found specialized modules A and C (**Fig. 2d**) and for LMAN specialized modules C,F,G (**Fig. 2c**). Unlike the module G specialization in HVC, both the female-E2 treated RA (module A,C) and LMAN (module C,F,G) module specializations were also present in females given the vehicle (**Fig. 2b**), making the observed RA and LMAN specializations unlikely to be causative for the gain of vocal learning in females after E2 treatment.

Telencephalic sexually responsive gene module

We next examined MEG associations to sex differences and E2 responses within the song system (**Fig. 2d**) and surrounding brain regions (**Fig. 2e**). Sex differences were compared between vehicle treated males and females. Here we found that module G was expressed higher in the vocal learning capable male HVC than the non-vocal learning capable female HVC. Module G similarly responded to sex within Area X, however since Area X is absent in vehicle treated females this comparison likely reflects the sexually dimorphic existence of Area X rather than a sex specific specialization. Module C expression was significantly higher in both female LMAN and RA relative to their male counterparts. In contrast, we found that gene module E was sexually dimorphic throughout the telencephalon, both in song nuclei and surrounding regions. Specifically, the eigengene for module E was highly expressed in males, and lowly expressed in females.

Gene modules enriched on specific chromosomes

We then determined whether any modules were enriched in genes from specific chromosomes. After bootstrapping the mapping between genes and modules to approximate a multiple test corrected upper-bound confidence interval for each chromosome-module pairing, we found genes in 8 of the 12 modules were enriched on specific chromosomes (**Fig. 3a**). The most striking effect observed was the enrichment of module E genes on the Z and W sex chromosomes, with nearly all W expressed genes and ~⅔ of Z expressed genes being members of module E. This explains why the eigengene of module E consistently exhibited higher expression in the male telencephalon regardless of treatment (**Fig. 2d,e**). When considering gene membership in module E as a continuous variable over the whole transcriptome by correlating each gene to the module's eigengene, we found that levels of Z and W transcripts were anticorrelated within this module (**Fig. 3b**), indicating that Z chromosome transcripts were generally depleted in female brains with W chromosome expression. This broad reduction in Z chromosome transcript abundance is consistent with the finding that diploid Z chromosomes do not inactivate in male birds to compensate for gene dosage, unlike the X inactivation seen in female mammals[28,33].

Other significant mappings included the largest module, module A, a component of RA's specialization, being slightly enriched across several macro-chromosomes (Chr1, 2, 3, 4, and 10). Interestingly, *FOXP2*, a language associated gene in humans[34] and vocal learning associated gene in songbirds[35] appeared to be a potent driver of module A, correlating with MEG-A at $r^2 = 0.92$. Conversely, module D, which did not associate with any experimental variables in our data, was enriched across micro-chromosomes (Chr22, 25, 27, 28, 29, and "other"). Module B, a component of the male HVC specialization, was enriched on chromosome 1A; module B also showed enrichment for genes with convergent specialized expression between zebra finch HVC and layer 2/3 neurons of the human laryngeal motor cortex[14] (**Fig. 2g**). No modules were enriched for the gene set convergently regulated between zebra finch RA and layer 5/6 neurons of the human laryngeal motor cortex (**Fig. 2h**). Module F, a component in HVC, LMAN, and Area X specializations, was enriched on chromosome 2. In addition to module D, both modules J and K were enriched on "other", a category that includes newly identified zebra finch micro-chromosomes[36]. These findings indicate a remarkable association between the structure of song nuclei gene expression networks, measured in an unbiased way with WGCNA, and the chromosomal structure of the genome.

Interactions between vocal learning and sex chromosome modules

As it is unlikely that the gene modules act independently of each other, we sought putative interacting genes between two modules of interest, the HVC vocal learning responsive module G and the sex chromosome dominated module E. To do this, we went beyond the binary, in-or-out module membership used to initially calculate MEGs and defined continuous module membership as the correlation (r^2) between a given gene's expression and any given MEG[31]. This allowed us to quantify the extent to which any gene is associated with any of the modules. Looking across all assigned genes for our modules of interest, we identified two outlier genes, *PDE8B* and *HABP4*, which were the E most gene assigned to module G and the G most gene assigned to module E, respectively (**Fig. 2f**). Both of these genes were found on the Z chromosome. *PDE8B* catalyzes the hydrolysis of the second messenger cAMP and mutations to the gene cause an autosomal dominant form of striatal degeneration in humans[37]. *HABP4* is an RNA binding protein, known to repress the expression and subsequent DNA binding of *MEF2C*[38], a Z chromosome transcription factor which has undergone a dramatic accelerated evolution in songbirds[15] and whose repression by FOXP2 is critical for cortico-striatal circuit formation in mice related to vocal behaviors[39].

Gene drivers of E2-treated female HVC specialization

To reduce the 344 genes in module G to the putative drivers of vocal learning in HVC and address our fourth question, we next examined the relationship of continuous membership in module G (correlation to MEG-G) to gene expression specialization in vocal learning capable HVC at the level of single genes. We did this separately for each of the four vocal learning comparisons discussed above: 1) male HVC specialization; 2) female E2 treated specialization; 3) HVC sexual dimorphism; 4) HVC E2 response (**Fig. 4a**). We defined genes of interest as having $r^2 \geq 0.5$ for both vocal learning specialization and module G membership in any single comparison (**Fig. 4b**). All such genes exhibited a positive correlation to vocal learning ability across sample sets (**Fig. 4b**), meaning that their expression was higher in vocal learning capable HVC relative to non-vocal learning controls.

We next intersected these four HVC specialized gene sets to generate the core genes list for module G involvement in E2 and sex dependent HVC development (**Fig. 4c**). We found 15 genes which strongly mark vocal learning capable HVC in all comparisons and also appear to drive module G expression (**Table 2, Fig. S3**). To address our fifth and final question, we examined the chromosomes of origin for these 15 core genes and found that four (*GHR*, *LRRC2*, *RGS7BP*, and *THBS4*) were on the Z sex chromsome (**Fig. 4d**). This > 4-fold enrichment of Z chromosome transcripts within the core gene set was statistically significant ($p = 0.0016$, upper-tailed hypergeometric test) against a background of module assigned genes. This result was statistically significant regardless of the background gene set when using all genes (~21,000) or only module G members (344). These four Z chromosome module G genes in HVC all exhibited > 50% reduced expression in vehicle-treated female HVC relative males, as expected for Z chromosome transcripts, but with increased HVC expression following E2 treatment (**Fig. S3**). This indicates that these transcripts are subject to additional, E2 sensitive transcriptional regulation that does not affect other genes from the Z chromosome in HVC.

Of the 15 core genes, several have been previously studied in the brains of other species which could inform their role in vocal learning systems. *THBS4* encodes a secreted extracellular-matrix glycoprotein necessary for appropriate neuronal migration in the mouse[40]and is elevated 6-fold in the human cortex compared with non-vocal learning primates[41]. Human *EDA2R* was recently identified as a top correlate of cognitive performance and brain size[42]; it was also found in a human GWAS study to correlate with circulating estrogen and testosterone levels[43]. Rare mutations in human *PHETA1* lead to Lowe oculocerebrorenal syndrome, the pathophysiology that includes seizures, mental retardation, and structural brain abnormalities[44,45]. *SIX2* is a homeobox domain containing transcription factor that governs early brain and craniofacial development, and provides neuroprotection from dopamine injury[46,47]. Perhaps one of the most relevant genes to atrophy of HVC in females, is the growth hormone receptor (*GHR*), which encodes a transmembrane receptor whose activation controls cell division[48]. The GHR ligand, growth hormone, has interestingly been shown to be duplicated and undergoing accelerated evolution in the genomes of songbirds, including the zebra finch[49]. Findings in these gene sets indicate that without estrogen treatment and induction of module G expression, females prevent the HVC expansion through a sex chromosome mechanism that blocks function of growth and neuronal migration promoting genes.

Implications for sexually dimorphic vocal learning

How did E2 treatment produce the transcriptomic effects we observed with module G in female HVC and what are the implications regarding sexually dimorphic vocal learning in the zebra finch? Presumably, this process begins with increased estrogen receptor (ER) activity within HVC cells after being provided surplus activating ligand[50], which is followed by altered transcription of ER targets in the genome. This initial transcriptional loading of the system is then processed by gene regulatory networks within each cell, spreading in effect through the transcriptome. It is possible that differential MEG-G expression arose purely from traditional gene regulatory networks, where transcription factors form complex, elaborate feedback networks with themselves, each other, and the genes they regulate. However, this framework fails to explain why module G transcriptional specializaton in vocal learning HVC was enriched for transcripts from a single chromosome; the Z sex chromosome with halved copy number in females.

We propose that the Z chromosome gene complex identified here may be in part controlled by a *cis*-acting gene regulatory element on the Z chromosome, activable downstream of ER activity. The genes regulated by this element may include necessary components in a proliferative transcriptional program, represented in our data as the enrichment of Z chromosome genes in the drivers of module G expression in HVC. The module G proliferative program could be specialized to developing male HVC by the expression of patterning genes, such as *SIX2* early in development, and plausibly maintained through persistent GHR signaling. We propose this cis-acting regulator is normally silent during female HVC development, a silence enforced by a combination of insufficient Z chromosome dosage, the expression of W chromosome transcripts, and sex differences in hormone levels. In the subsequent absence of gene products such as GHR, module G does not specialize in developing female HVC leading to a less proliferative female HVC late in development. We term this hypothesis the "Z chromosome *cis*-regulatory song model" of sexually dimorphic zebra finch vocal learning (**Fig. 5**).

Our results challenge the traditional notion that hormonal signaling organizes the brain to produce sexually dimporhic behaviors independent of neuronal sex chromosome content[51,52]. Instead, our data indicated that sexual dimorphism in zebra finch song learning ability was established by the interaction of sex hormone signaling and sex chromsome gene expression within the HVC lineage during development. These findings are thematically similar to work in the Four Core Genotypes mouse line[53], where chromosomal and gonadal sex are sepparable due to translocation of the sex determining *Sry* gene. In these mice, sex chromosome compostion has been shown to affect sexaully dimorphic circuit anatomy, behaviors, and gene expression[53–56], though the sex chromosome genes responsible remain unknown in all instances. By leveraging the discrete nature of the song system, the female zebra finch response to E2, and modern genomic technologies, here we were able to identify specific candidate genes of vocal learning loss from the Z sex chromosome and describe the sexually dimorphic, HVC specialized gene network which they participate in.

In conclusion, we show that E2 treatment in females induces a subset of gene expression specializations seen in male HVC and that this specialization is disproportionately influenced by the expression of Z chromosome transcripts. We hypothesize that these genes are regulated by a Z chromosome *cis*-regulatory element and are necessary for continued HVC proliferation late in development. Additional experiments manipulating these genes to both mimic and prevent the action of E2 in female zebra finches are needed to test this hypothesis, as are comparisons of zebra finch Z chromosome structural and epigenomic profiles across sex, hormonal manipulation, and development.

Acknowledgments

We would like to thank G. Gedman, G. Formenti, C. Gilbert, L. Cantin, C. Vargas and J. Manley for conversations during analysis and visualization. We'd also like to thank A. Vaziri and T. Nöbauer for providing the computing infrastructure used throughout. This work is a memorial to M. Konishi (1933-2020) whose work on this topic has influenced us all greatly.

Author contributions

MD performed all analyses and wrote the paper, HNC gathered all data and edited the paper, HM and EDJ edited the paper.

Materials & Correspondence

Correspondence to Erich D. Jarvis.

Competing interests

We have no competing interests.

Materials and Methods

Animal Handling and Sample Preparation:

The samples used in the present analysis are the E2 or vehicle treated subset of a previously published RNAseq dataset[26]. We briefly re-describe our methodology here.

E2 (Sigma E1024-1G) was dissolved in DMSO (100mg/mL) and then diluted in olive oil (1mg/mL). 30-50uL of E2 sample or DMSO only vehicle was applied to the flank of male and female zebra finches daily from PHD0-PHD14 and on alternating days from PHD15-30 (n=3 per sex-treatment combination). We have previously shown that this treatment program is sufficient to induce song system masculinization in E2 treated female zebra finches[26].

On PHD30, animals were sacrificed following one hour of dark isolation. Animals were anesthetized by isoflurane inhalation and rapidly decapitated. Brain hemispheres were dissected, embedded in OCT, and flash frozen in an ethanol and dry ice slurry. Sections were taken from the right hemisphere coronally at 14um onto polyethylene naphthalate (PEN) membrane slides for RNA isolation and adjacent sections taken on charged glass slides for histology or in-situ hybridizations. From the PEN membrane slides, song nuclei and surrounding control regions were laser capture microdissected (LCM) using an ArcturusXT LCM system (Nikon) guided by a Nissl stained tissue series for each animal.

RNA was extracted from the LCM isolated tissue samples using the Arcturus Picopure kit (Applied Biosystems KIT0204) following manufacturer's instructions. RNA quality was assessed using an Agilent 2100 bio-analyzer and the RNA 6000 pico kit (Agilent 5067- 1513). Next, cDNA was synthesized using the SMART-Seq v4 Ultra Low input RNA Kit (Takara 634892). Sequencing libraries were made with the NEBNext Ultra II DNA Library Prep kit (New England Biolabs E7645L) and cleaned-up using SPRIselect beads (Beck-man Coulter B23317). Libraries were sequenced by Novogene Co., Ltd. on the Novaseq 6000 platform (Illumina) and S4 flow cells resulting in 150bp paired-end reads.

RNAseq read mapping and quality control:

RNAseq reads were first trimmed to remove adapters and low quality base calls using Trimmomatic[57] and then mapped to a high-quality Vertebrate Genomes Project (VGP) female zebra finch nuclear genome (bTaeGut2.pat.W.v2, GCF_008822105.2)[5] using STAR (v2.7.1)[58]. These mappings were then tallied at the level of genes using Rsubread::featureCounts (R-3.6.1) and then counts normalized to fragments per kilobase of transcript per million mapped reads[59]. Read based quality control was performed with FastQC (Babraham Bioinformatics) with reports prepared by MultiQC (Python-3.5.5)[60]. This workflow was automated by the CountMatrix pipeline (https://github.com/mattisabrat/CountMatrix/). We next removed two outlier samples (one male vehicle HVC and one female vehicle RA) based upon hierarchical clustering of the sample space before computing gene to gene correlations (**Fig. S1**).

Gene Module Identification:

All remaining analyses were completed in R-3.7.1 and Bioconductor-v19. Data was wrangled in the tidyverse, and custom visualizations produced with ggplot, RColorBrewer, and ggpubr[61,62]. Unsigned topological overlaps between genes (gene-to-gene correlations) were calculated in a single block with WGNCA::blockwiseModules, and then module assignment was performed using WGCNA::recutBlockwiseTrees specifying a minimum module size of 100 genes[31].

Module association to vocal learning:

Module eigengenes (MEGs) from each module were correlated with song system membership/vocal learning capability and the statistical significance of each correlation assessed using WGCNA::corPvalueStudent. This was done in the following sample subsets by node: male samples; female samples; female samples treated with vehicle; song system components from either sex treated with vehicle; and female song system components of either treatment. There are four comparisons of vocal learning groups to be drawn within these subsets; the male song system against the surrounds; the female E2 treated song system against all other female samples; the vehicle treated male song system against the vehicle treated female song system; and the female E2 treated song system against the female vehicle treated song system. The comparison of the vehicle treated female song system against the surrounds (correlation to song system variable) was also performed. In cases where there were outliers on both sides of a comparison originating from the same animal, we performed post-hoc two-tailed Wilcoxon rank sum tests using wilcox.test to associate the rank order of MEG expression to the variable at play for those comparisons.

Module gene ontology and convergent vocal learning gene expression signature enrichment:

Module assigned zebra finch genes were mapped to their 1:1 human orthologs where possible, dropping unmapped or multi-mapped genes, using orthofindR::getOrthos (https://github.com/ggedman/orthofindR) which wraps Ensembl's biomaRt. We tested for enriched human gene ontology biological process terms within the human orthologs of each module using generally applicable gene set enrichment (GAGE) implemented in gage::gage[63]. To determine if the genes previously shown as convergently differentially expressed in zebra finch RA or HVC and human dorsal laryngeal motor cortex mapped to specific modules, we treated these two gene lists identically to GO terms and tested for their significant enrichment across the human orthologs of each module also using GAGE.

Module enrichment on chromosomes:

To associate modules to chromosomes, we bootstrapped Bonferonni corrected upper bound 95% confidence intervals for the enrichment of each chromosome-module pairing by randomizing the mapping of genes to modules 25k times and calculating the enrichments observed on each chromosome from each module in each randomization to approximate the null distributions.

Identification of core module G genes in HVC and Z chromosome enrichment:

We defined genes of interest as the subset of significantly ($p < 0.05$) vocal learning capability correlated genes in HVC whose expression correlated to module eigengene G (MEG-G) across the dataset at $r^2 \geq 0.5$ and to vocal learning at $r^2 \geq 0.5$ in at least one of the four vocal learning comparisons in HVC, again calculated using WGCNA::corPvalueStudent. These comparisons were: all male HVC samples against all male DN samples; female E2 treated HVC against all other female samples at the node, including vehicle treated HVC; Vehicle treated male HVC against vehicle treated female HVC; and E2 treated female HVC against vehicle treated female HVC. We defined core genes as those meeting this criteria for all four HVC vocal learning comparisons. We tested the statistical significance of Z chromosome enrichment in this core gene list with an upper-tailed hypergeometric test, implemented in phyper, where each core gene is a sampling event without replacement from module assigned genes.

References

1. Nottebohm, F. & Arnold, A. P. Sexual dimorphism in vocal control areas of the songbird brain. *Science* **194**, 211–213 (1976).
2. Gurney, M. E. & Konishi, M. Hormone-induced sexual differentiation of brain and behavior in zebra finches. *Science* **208**, 1380–1383 (1980).
3. Simpson, H. B. & Vicario, D. S. Early estrogen treatment of female zebra finches masculinizes the brain pathway for learned vocalizations. *J. Neurobiol.* **22**, 777–793 (1991).
4. Simpson, H. B. & Vicario, D. S. Early estrogen treatment alone causes female zebra finches to produce learned, male-like vocalizations. *J. Neurobiol.* **22**, 755–776 (1991).
5. Rhie, A. *et al.* Towards complete and error-free genome assemblies of all vertebrate species. *Nature* **592**, 737–746 (2021).
6. Jarvis, E. D. Evolution of vocal learning and spoken language. *Science* **366**, 50–54 (2019).
7. Kumar, S. & Hedges, S. B. A molecular timescale for vertebrate evolution. *Nature* **392**, 917–920 (1998).
8. van Tuinen, M. & Hadly, E. A. Error in estimation of rate and time inferred from the early amniote fossil record and avian molecular clocks. *J. Mol. Evol.* **59**, 267–276 (2004).
9. Bolhuis, J. J., Okanoya, K. & Scharff, C. Twitter evolution: converging mechanisms in birdsong and human speech. *Nat. Rev. Neurosci.* **11**, 747–759 (2010).
10. Doupe, A. J. & Kuhl, P. K. Birdsong and human speech: common themes and mechanisms. *Annu. Rev. Neurosci.* **22**, 567–631 (1999).
11. Mooney, R. Neural mechanisms for learned birdsong. *Learn. Mem.* **16**, 655–669 (2009).
12. Pfenning, A. R. *et al.* Convergent transcriptional specializations in the brains of humans and song-learning birds. *Science* **346**, 1256846 (2014).
13. Feenders, G. *et al.* Molecular mapping of movement-associated areas in the avian brain: a motor theory for vocal learning origin. *PLoS One* **3**, e1768 (2008).
14. Gedman, G. Songbird brain organization and its molecular convergence with humans for vocal imitation learning. (Rockefeller University, 2021).
15. James A. Cahill, Joel Armstrong, Alden Deran, Carolyn J. Khoury, Benedict Paten, David Haussler, Erich D. Jarvis. Positive selection in non-coding genomic regions of vocal learning birds is associated with genes implicated in vocal learning and speech functions in humans. *Genome Res.*
16. The zebra finch: a synthesis of field and laboratory studies. *Choice* **34**, 34–5095–34–5095 (1997).
17. Bottjer, S. W., Glaessner, S. L. & Arnold, A. P. Ontogeny of brain nuclei controlling song learning and behavior in zebra finches. *J. Neurosci.* **5**, 1556–1562 (1985).
18. Garcia-Calero, E. & Scharff, C. Calbindin expression in developing striatum of zebra finches and its relation to the formation of area X. *Journal of Comparative Neurology* vol. 521 326–341 (2013).
19. Konishi, M. & Akutagawa, E. Neuronal growth, atrophy and death in a sexually dimorphic song nucleus in the zebra finch brain. *Nature* **315**, 145–147 (1985).
20. Nixdorf-Bergweiler, B. E. Divergent and parallel development in volume sizes of telencephalic song nuclei in male and female zebra finches. *J. Comp. Neurol.* **375**, 445–456 (1996).
21. Nordeen, E. J. & Nordeen, K. W. Sex and regional differences in the incorporation of neurons born during song learning in zebra finches. *J. Neurosci.* **8**, 2869–2874 (1988).
22. Shaughnessy, D. W., Hyson, R. L., Bertram, R., Wu, W. & Johnson, F. Female zebra finches do not sing yet share neural pathways necessary for singing in males. *J. Comp. Neurol.* **527**, 843–855 (2019).
23. Holloway, C. C. & Clayton, D. F. Estrogen synthesis in the male brain triggers development of the avian song control pathway in vitro. *Nat. Neurosci.* **4**, 170–175 (2001).
24. Mooney, R. & Rao, M. Waiting periods versus early innervation: the development of axonal connections in the zebra finch song system. *J. Neurosci.* **14**, 6532–6543 (1994).

25. Gurney, M. E. Behavioral correlates of sexual differentiation in the zebra finch song system. *Brain Research* vol. 231 153–172 (1982).
26. Choe, H. N. *et al.* Estrogen and sex-dependent loss of the vocal learning system in female zebra finches. *Horm. Behav.* **129**, 104911 (2021).
27. Herrmann, K. & Arnold, A. P. Lesions of HVc block the developmental masculinizing effects of estradiol in the female zebra finch song system. *J. Neurobiol.* **22**, 29–39 (1991).
28. Agate, R. J. *et al.* Neural, not gonadal, origin of brain sex differences in a gynandromorphic finch. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 4873–4878 (2003).
29. Gianaroli, L., Magli, M. C. & Ferraretti, A. P. Sex Chromosomes. in *Brenner's Encyclopedia of Genetics (Second Edition)* (eds. Maloy, S. & Hughes, K.) 397–400 (Academic Press, 2013).
30. Konishi, M. & Akutagawa, E. A critical period for estrogen action on neurons of the song control system in the zebra finch. *Proc. Natl. Acad. Sci. U. S. A.* **85**, 7006–7007 (1988).
31. Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 559 (2008).
32. Gedman, G. *et al.* As above, so below: Whole transcriptome profiling demonstrates strong molecular similarities between avian dorsal and ventral pallial subdivisions. *J. Comp. Neurol.* (2021) doi:10.1002/cne.25159.
33. Itoh, Y. *et al.* Dosage compensation is less effective in birds than in mammals. *J. Biol.* **6**, 2 (2007).
34. Lai, C. S., Fisher, S. E., Hurst, J. A., Vargha-Khadem, F. & Monaco, A. P. A forkhead-domain gene is mutated in a severe speech and language disorder. *Nature* **413**, 519–523 (2001).
35. Teramitsu, I. & White, S. A. FoxP2 regulation during undirected singing in adult songbirds. *J. Neurosci.* **26**, 7390–7394 (2006).
36. Kim, J. *et al.* False gene and chromosome losses affected by assembly and sequence errors. *bioRxiv* 2021.04.09.438906 (2021) doi:10.1101/2021.04.09.438906.
37. Appenzeller, S. *et al.* Autosomal-Dominant Striatal Degeneration Is Caused by a Mutation in the Phosphodiesterase 8B Gene. *The American Journal of Human Genetics* vol. 86 83–87 (2010).
38. Kobarg, C. B., Kobarg, J., Crosara-Alberto, D. P., Theizen, T. H. & Franchini, K. G. MEF2C DNA-binding activity is inhibited through its interaction with the regulatory protein Ki-1/57. *FEBS Lett.* **579**, 2615–2622 (2005).
39. Chen, Y.-C. *et al.* Foxp2 controls synaptic wiring of corticostriatal circuits and vocal communication by opposing Mef2c. *Nat. Neurosci.* **19**, 1513–1522 (2016).
40. Girard, F., Eichenberger, S. & Celio, M. R. Thrombospondin 4 deficiency in mouse impairs neuronal migration in the early postnatal and adult brain. *Mol. Cell. Neurosci.* **61**, 176–186 (2014).
41. Cáceres, M., Suwyn, C., Maddox, M., Thomas, J. W. & Preuss, T. M. Increased cortical expression of two synaptogenic thrombospondins in human brain evolution. *Cereb. Cortex* **17**, 2312–2321 (2007).
42. Harris, S. E. *et al.* Neurology-related protein biomarkers are associated with cognitive ability and brain volume in older age. *Nat. Commun.* **11**, 800 (2020).
43. Ruth, K. S. *et al.* Using human genetics to understand the disease impacts of testosterone in men and women. *Nat. Med.* **26**, 252–258 (2020).
44. Kornfeld, M., Synder, R. D., MacGee, J. & Appenzeller, O. The oculo-cerebral-renal syndrome of Lowe. *Arch. Neurol.* **32**, 103–107 (1975).
45. Ates, K. M. *et al.* Deficiency in the endocytic adaptor proteins PHETA1/2 impairs renal and craniofacial development. *Dis. Model. Mech.* **13**, (2020).
46. Garcez, R. C., Le Douarin, N. M. & Creuzet, S. E. Combinatorial activity of Six1-2-4 genes in cephalic neural crest cells controls craniofacial and brain development. *Cell. Mol. Life Sci.* **71**, 2149–2164 (2014).
47. Gao, J. *et al.* Transcription factor Six2 mediates the protection of GDNF on 6-OHDA lesioned dopaminergic neurons by regulating Smurf1 expression. *Cell Death Dis.* **7**, e2217 (2016).

48. Dehkhoda, F., Lee, C. M. M., Medina, J. & Brooks, A. J. The Growth Hormone Receptor: Mechanism of Receptor Activation, Cell Signaling, and Physiological Aspects. *Front. Endocrinol.* **9**, 35 (2018).
49. Yuri, T., Kimball, R. T., Braun, E. L. & Braun, M. J. Duplication of accelerated evolution and growth hormone gene in passerine birds. *Mol. Biol. Evol.* **25**, 352–361 (2008).
50. Frankl-Vilches, C. & Gahr, M. Androgen and estrogen sensitivity of bird song: a comparative view on gene regulatory levels. *J. Comp. Physiol. A Neuroethol. Sens. Neural Behav. Physiol.* **204**, 113–126 (2018).
51. Phoenix, C. H., Goy, R. W., Gerall, A. A. & Young, W. C. Organizing action of prenatally administered testosterone propionate on the tissues mediating mating behavior in the female guinea pig. *Endocrinology* **65**, 369–382 (1959).
52. Arnold, A. P. The organizational-activational hypothesis as the foundation for a unified theory of sexual differentiation of all mammalian tissues. *Horm. Behav.* **55**, 570–578 (2009).
53. De Vries, G. J. *et al.* A model system for study of sex chromosome effects on sexually dimorphic neural and behavioral traits. *J. Neurosci.* **22**, 9005–9014 (2002).
54. Chen, X., Grisham, W. & Arnold, A. P. X chromosome number causes sex differences in gene expression in adult mouse striatum. *Eur. J. Neurosci.* **29**, 768–776 (2009).
55. McPhie-Lalmansingh, A. A., Tejada, L. D., Weaver, J. L. & Rissman, E. F. Sex chromosome complement affects social interactions in mice. *Horm. Behav.* **54**, 565–570 (2008).
56. Cox, K. H. & Rissman, E. F. Sex differences in juvenile mouse social behavior are influenced by sex chromosomes and social context. *Genes Brain Behav.* **10**, 465–472 (2011).
57. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
58. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
59. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* vol. 30 923–930 (2014).
60. Ewels, P., Magnusson, M., Lundin, S. & Käller, M. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* **32**, 3047–3048 (2016).
61. Wickham, H. *et al.* Welcome to the tidyverse. *J. Open Source Softw.* **4**, 1686 (2019).
62. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*. (Springer Science & Business Media, 2009).
63. Luo, W., Friedman, M. S., Shedden, K., Hankenson, K. D. & Woolf, P. J. GAGE: generally applicable gene set enrichment for pathway analysis. *BMC Bioinformatics* **10**, 161 (2009).

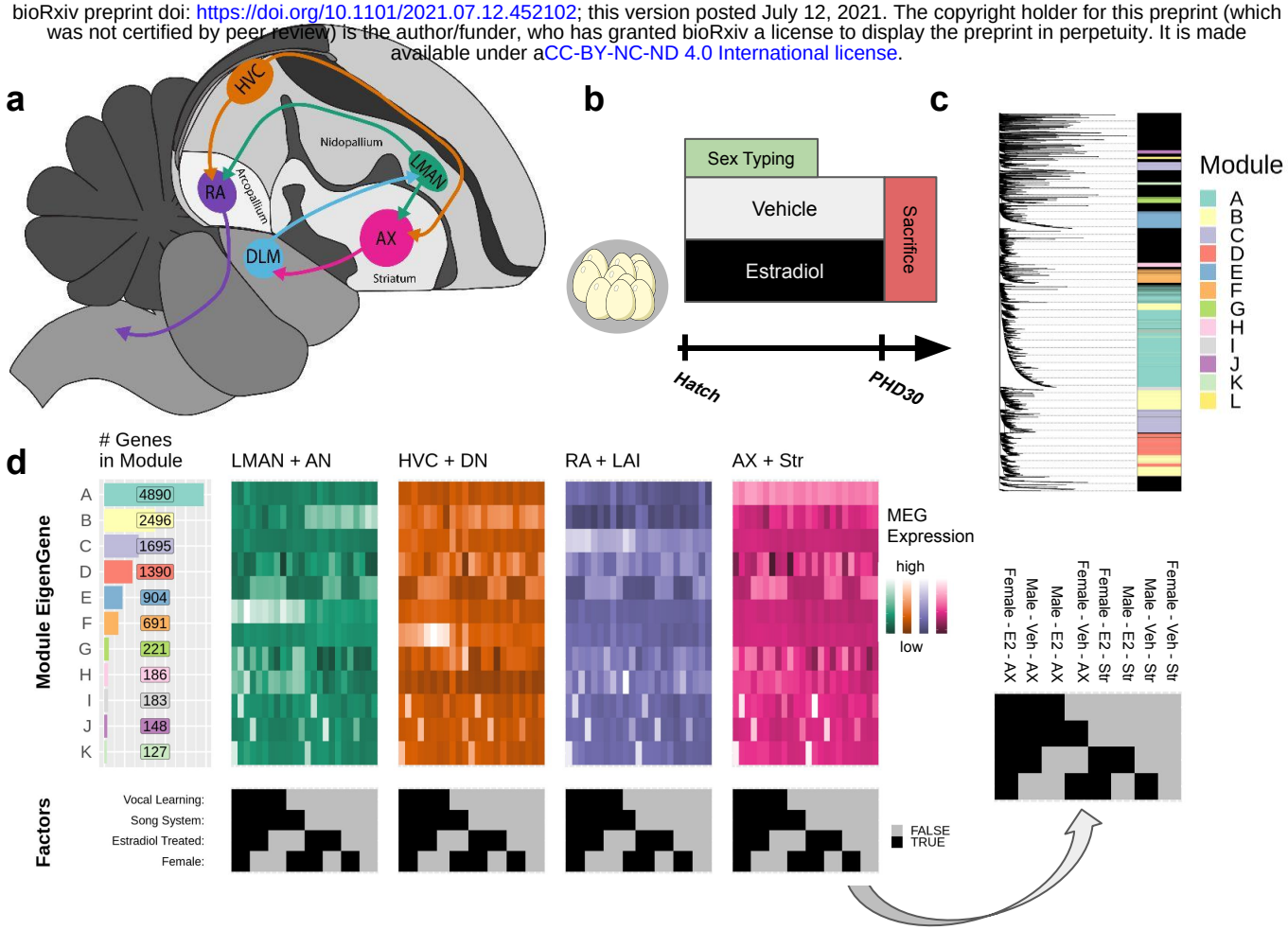


Figure 1 - Song system anatomy and experimental design. **a**, Diagram of song system connectivity within the adult male zebra finch brain. Area X connects back to LMAN through the non-vocal specific thalamic nucleus DLM. **b**, Experimental design. Animals were treated with E2 or a vehicle from hatch until sacrifice on PHD30. **c,** WGCNA and assignment of genes to modules. Left: Hierarchy computed over the transcriptome wide topological overlap matrix. Right: Rows are genes colored according to the assigned module, unassigned genes in black. **d,** MEG expression heatmaps by module size (left) aligned to traits of interest (bottom). Each row is an MEG and each sample is a column with samples grouped by neural circuit node into different colored subpanels.

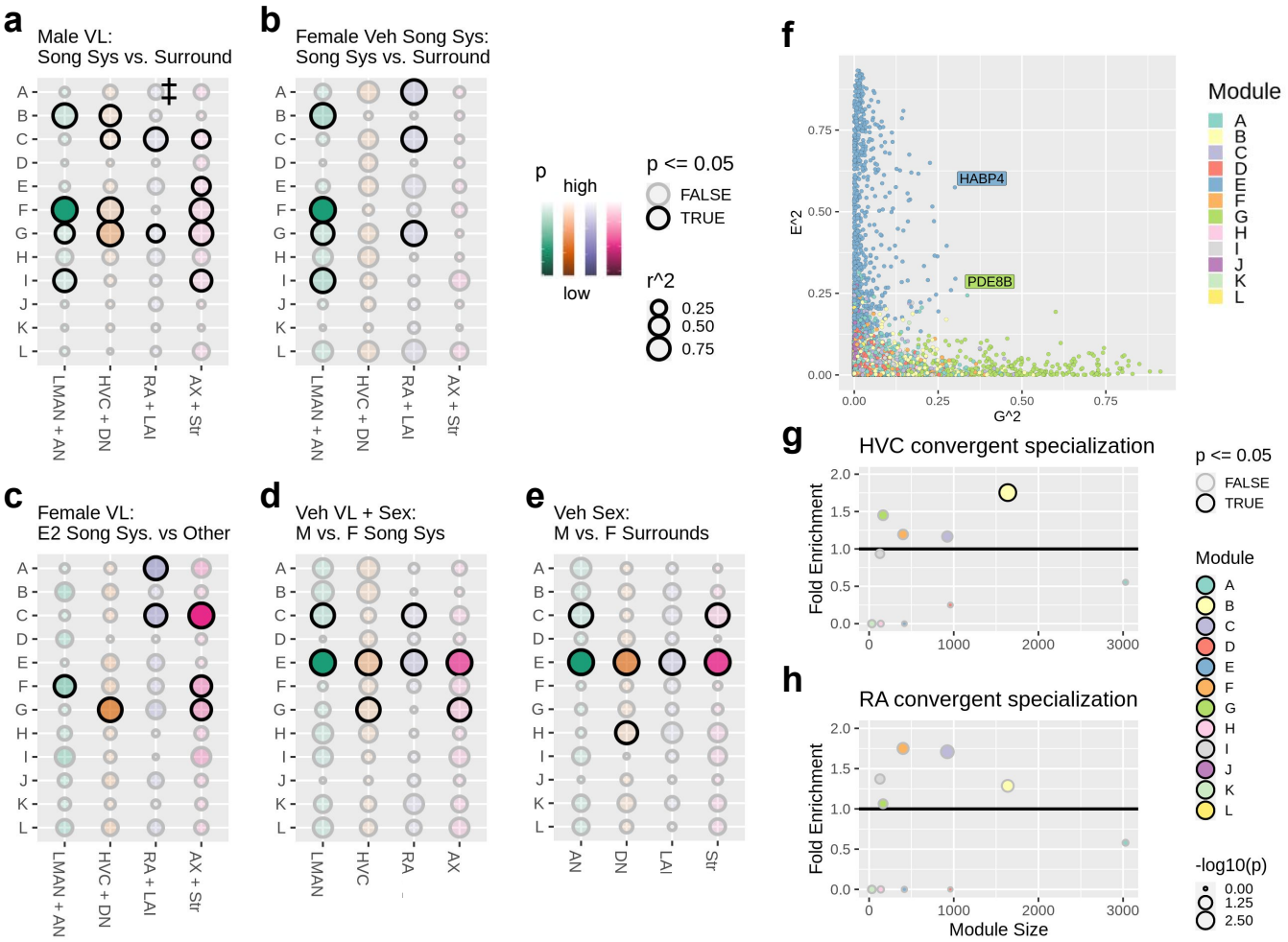
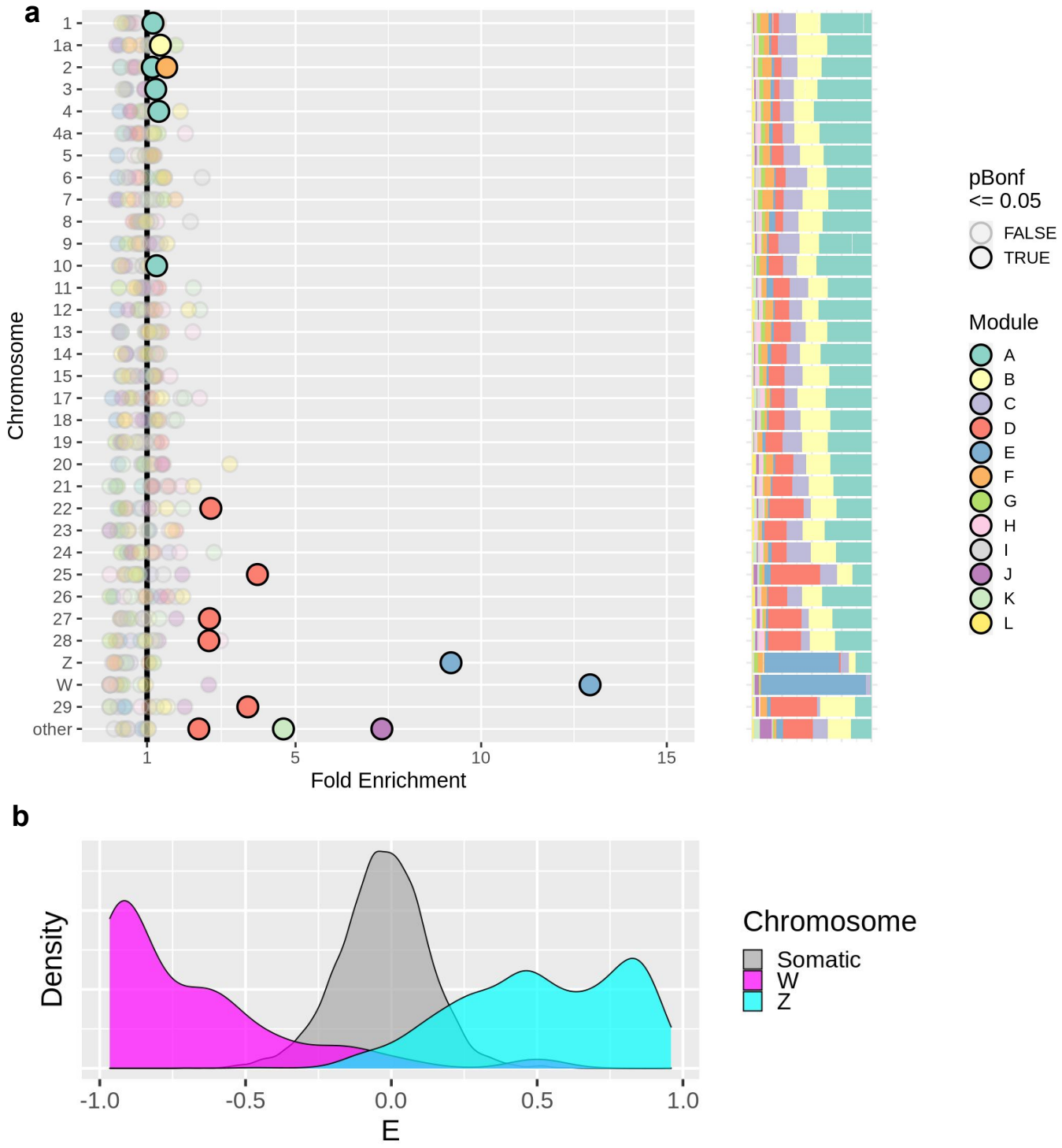


Figure 2 - Association of modules to specific variables. **a-e,** Results of statistical association between MEG expression and variables of interest. Plots show the associations between gene modules (rows) to **a,** male song system membership; **b,** vehicle treated female song system membership; **c,** female vocal learning after E2; **d,** sexual dimorphism within the song system; **e,** sexual dimorphism within the surrounding control regions. Each neural circuit node is considered separately (columns). Pearson correlation and Students t test, α=0.05. Where T test is inappropriate because of paired outliers, ‡ indicates p=0.06 by post hoc Wilcoxon rank sum test. **f,** Comparison of membership in module E (r^2 to MEG-E, y-axis) and module G (r^2 to MEG-G, x-axis) across module assigned genes. **g,** Enrichment of genes previously found to be convergently differentially expressed in the human laryngeal motor cortex and either HVC (top) or RA (bottom). Values above theblack line indicate above random chance.



528 **Figure 3 - Gene modules on chromosomes. a**, Fold enrichment of modules onto chromosomes
529 (left) and raw data, the proportion of genes from each chromosome assigned to each module (right).
530 Values to the right of the vertical black line indicate above random chance. Significance was
531 assessed using a bootstrapped test of observed enrichment against Bonferroni corrected 95% upper
532 bounds estimated for each module chromosome pairing based on 25,000 randomizations. **b**,
533 Distribution of continuous membership in module E across the module assigned transcriptome
534 (Pearson r to MEG-E: E) with sex chromosomes separated.

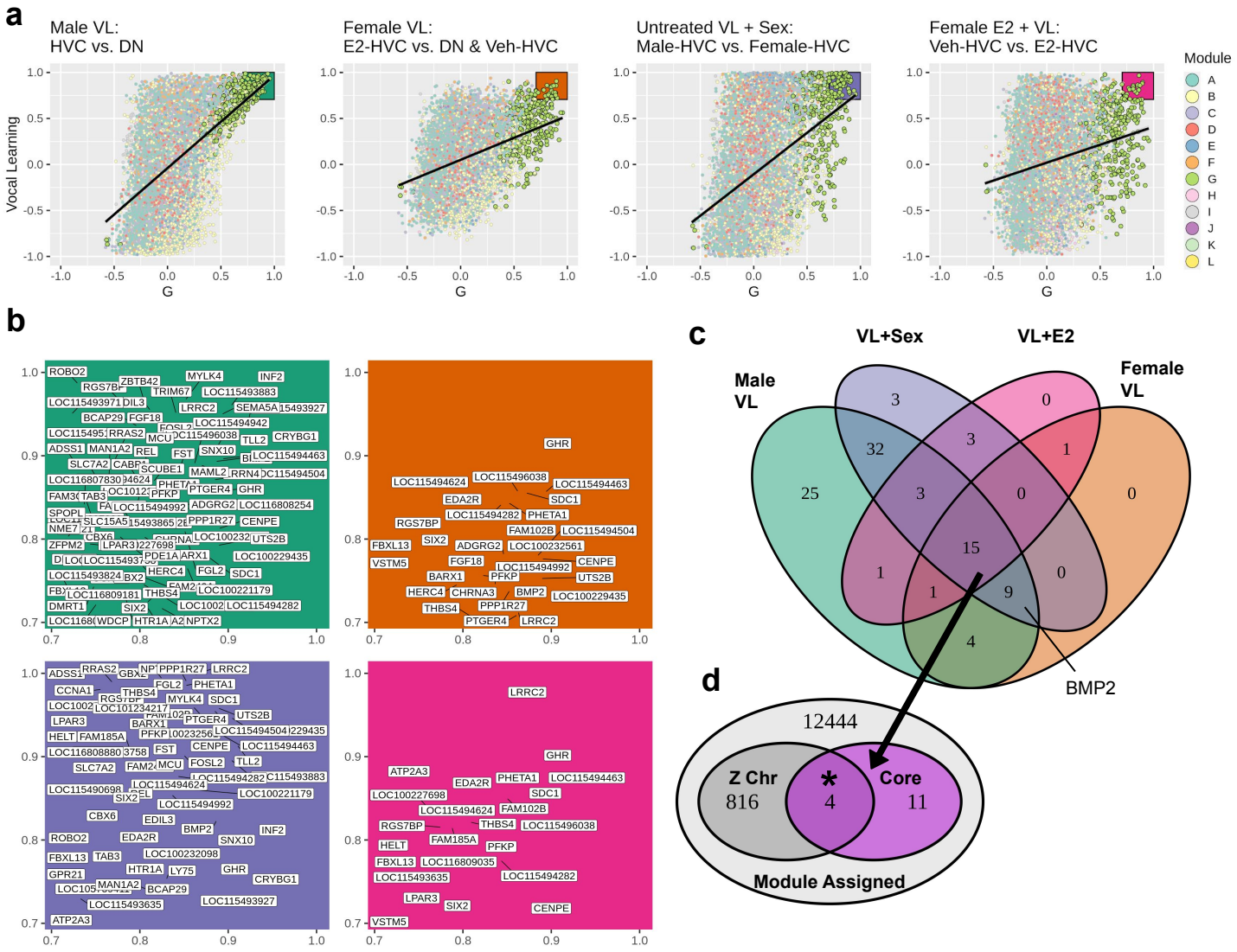
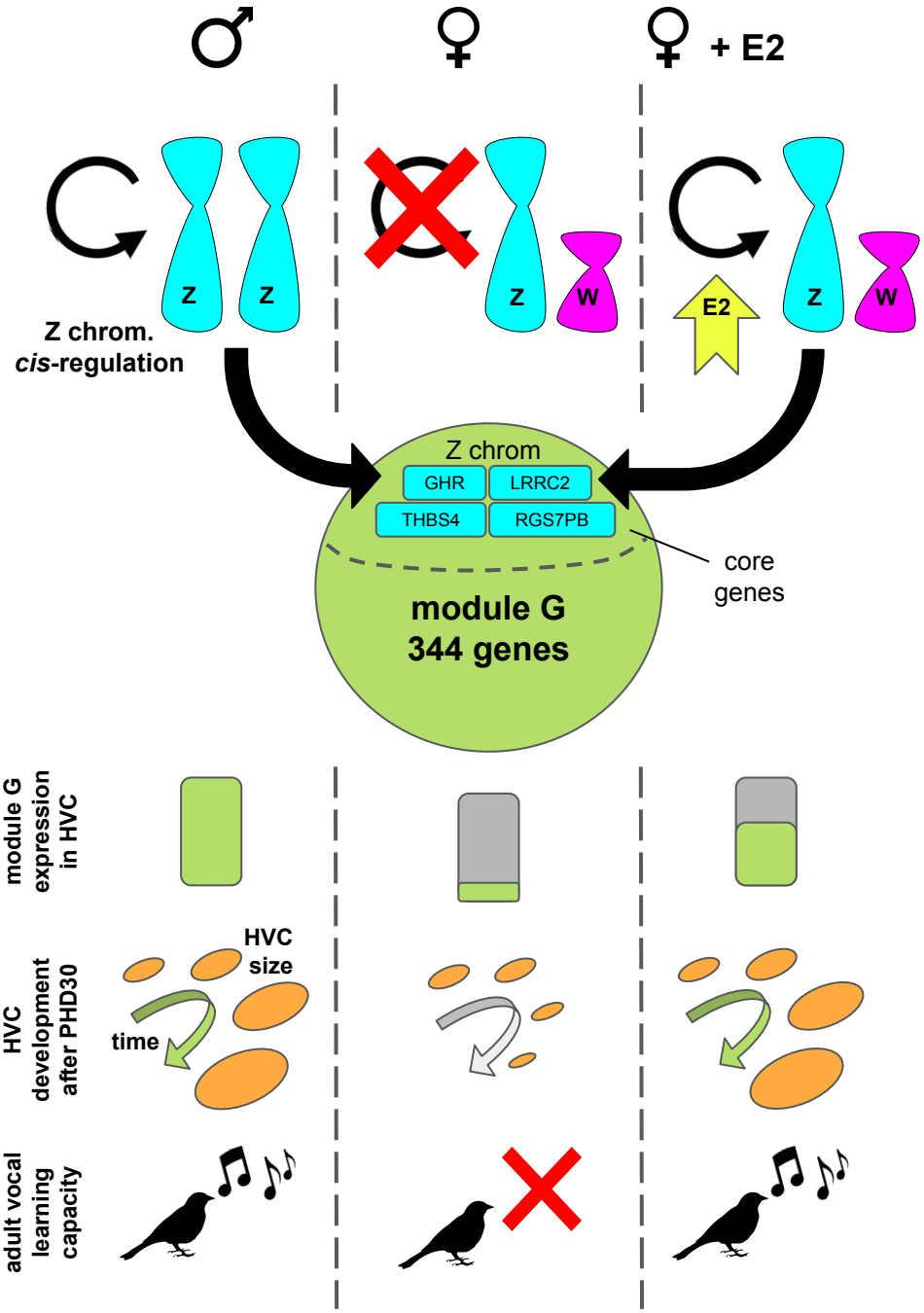


Figure 4 - Identification of core genes and association to the Z chromosome. a, Single gene continuous membership in module G (x-axis; Pearson r to MEG from module G) for all assigned genes vs correlation to vocal learning in masculine or masculinized HVC relative to samples from non-vocal learning females in each of the four comparisons (titles). Shaded area indicates gene of interest criteria for each comparison. **b,** Blowup of shaded regions in **a** showing genes of interest from each comparison. **c,** Identification of core genes by intersecting the four gene sets of interest. **d,** Enrichment of Z chromosome transcripts within the core genes. * indicates $p < 0.05$ by an upper-tailed hypergeometric test.



543 **Figure 5 - Z chromosome *cis*-regulatory hypothesis of sexually dimorphic zebra finch vocal**
544 **learning.** We propose that estradiol treatment in female zebra finches masculinizes song behavior
545 by overcoming insufficient Z sex chromosome dosage to activate a Z chromosome *cis*-acting
546 regulator in HVC which would otherwise be male specific. The genes controlled by this putative
547 *cis*-acting element are drivers of a larger proliferative genetic program which prevents HVC atrophy
548 and allows for its expansion late in development, sufficient for rudimentary vocal learning in females.

Module G: Significantly Enriched GO

GOterm	GOid	p
DNA-binding transcription factor activity, RNA polymerase II-specific	GO:0000981	0.012
cell differentiation	GO:0030154	0.013
DNA-binding transcription factor activity	GO:0003700	0.015
extracellular matrix structural constituent	GO:0005201	0.024
inflammatory response	GO:0006954	0.026
anatomical structure morphogenesis	GO:0009653	0.03
external side of plasma membrane	GO:0009897	0.036
extracellular space	GO:0005615	0.037
DNA-binding transcription activator activity, RNA polymerase II-specific	GO:0001228	0.04
cell-cell signaling	GO:0007267	0.041
positive regulation of multicellular organism growth	GO:0040018	0.042
transmembrane transporter activity	GO:0022857	0.043

Table 1 - Module G functional enrichment analysis. Significantly enriched GO terms within the 1:1 human orthologs of module G. Lists full GO term, GOid, and p value calculated by generally applicable gene set enrichment[47].

Module G: Core VL Genes

Z Chromosome	Other	cont.
GHR	CENPE	LOC115494624
LRRC2	EDA2R	PFKP
RGS7BP	FAM102B	PHETA1
THBS4	FBXL13	SDC1
	LOC115494282	SIX2
	LOC115494463	

Table 2 - Core genes of module G specialization to vocal learning HVC. Putative drivers of module G specialization to vocal learning capale HVC. Intersection of the four gene of interest sets (**Fig. 4b**), separating the significant enrichment of genes from the Z sex chromosome.

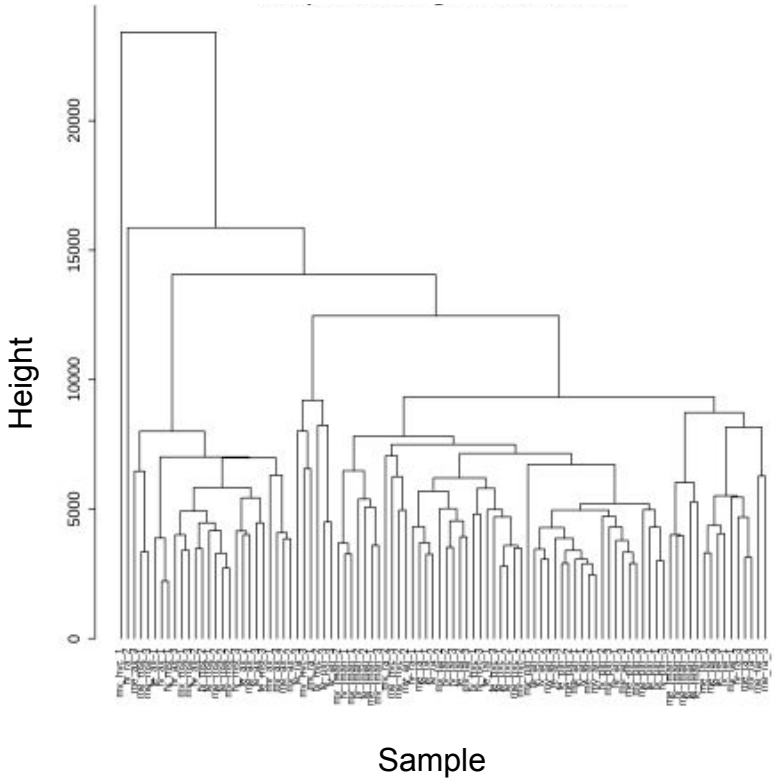


Figure S1 - Outlier sample detection by hierarchical clustering. Two samples (mv_hvc_1 and fe_ra_2, left most) form deep, single sample branches in the hierarchical clustering tree , indicative of outlier samples unlikely to fit the correlational structure of the larger dataset. Samples were removed prior to gene network construction and module detection.

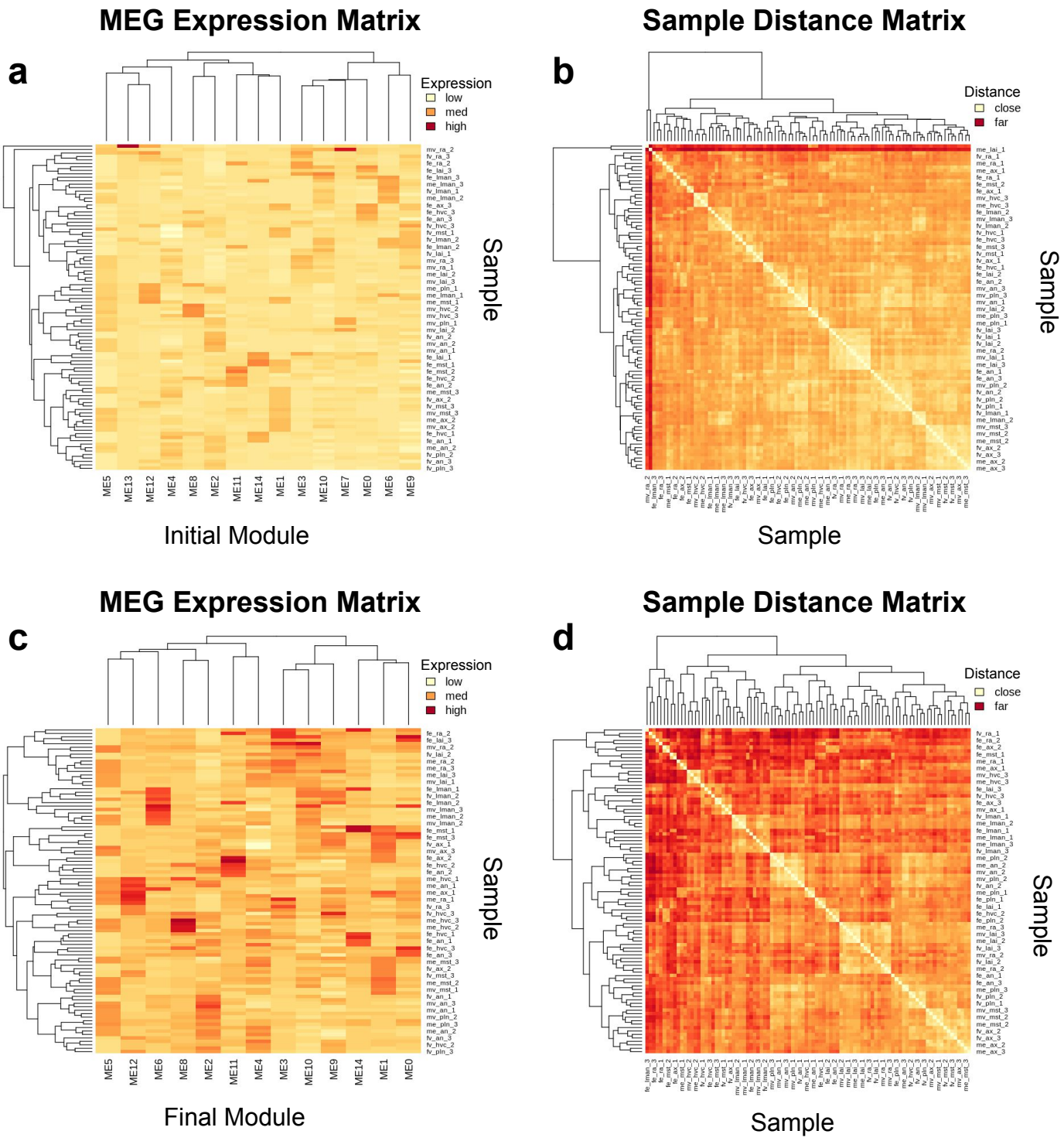
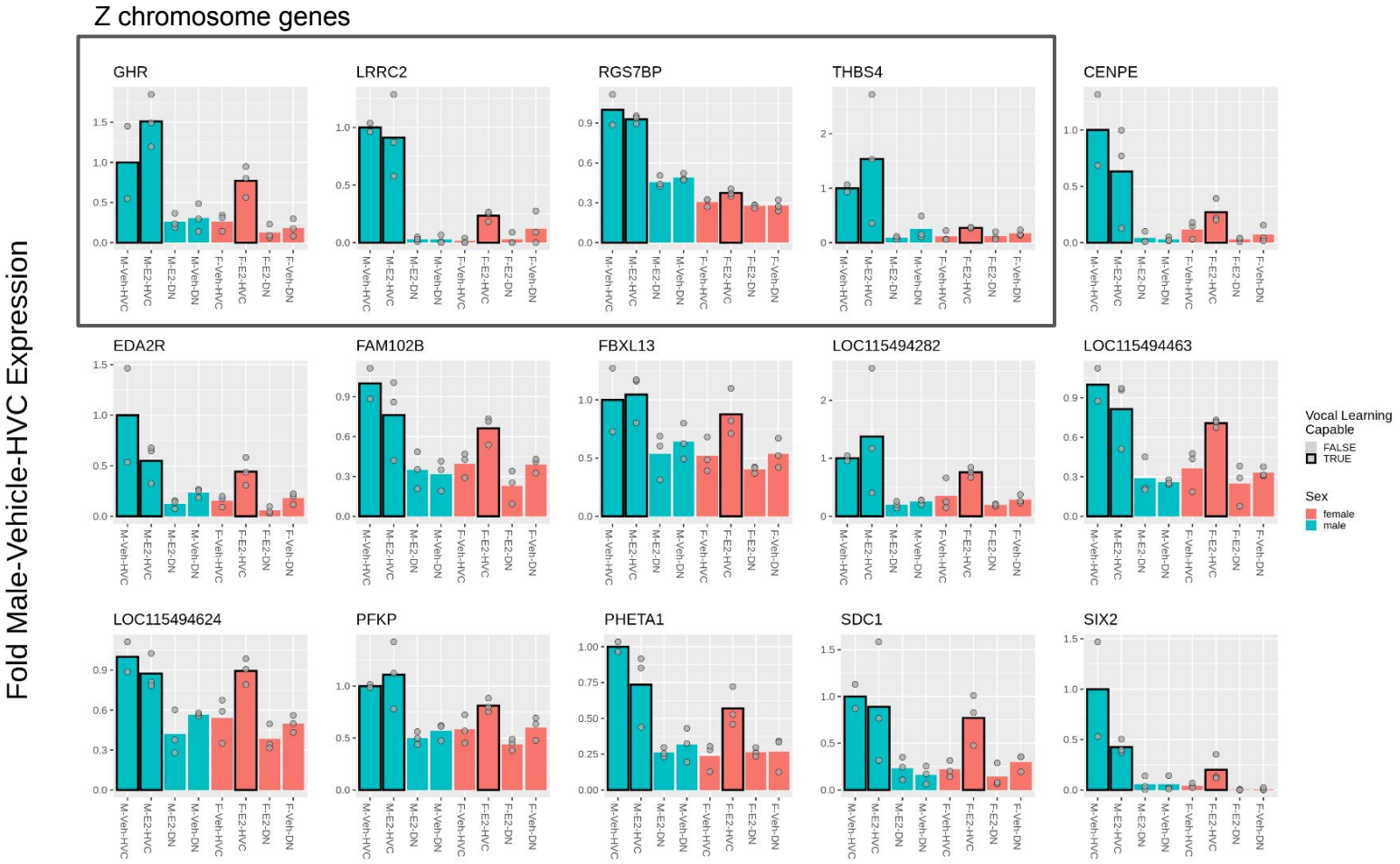


Figure S2 - Initial module overfitting to single samples. a, ME7 is highly expressed only in mv_ra_2 and ME13 is highly expressed only in fv_ra_3. **b,** This overfitting causes these samples to be deep outliers in the sample-sample distance matrix. **c,** Removing these module eigengenes from the set prevents these samples from behaving as outliers in the distance matrix in **d**. These overfit modules were removed prior to module lettering and statistical analysis.



564 **Figure S3 - Expression of module G core genes in HVC and surrounding dorsal nidopallium.**
565 Each of the 15 core genes show reduced expression in female HVC relative to the male with an
566 increase in expression in response to E2 treatment. This transcriptional response to E2 is not seen in
567 the surrounding DN.

Gene	G	Gene	G	Gene	G	Gene	G	Gene	G	Gene	G	Gene	G	Gene	G
CRYBG1	0.96	EDA2R	0.81	LOC105759411	0.74	LOC115491275	0.68	TWIST1	0.64	LOC115490816	0.58	ATG3	0.51	LOC116808453	0.43
INF2	0.94	LOC115494624	0.81	LOC116807830	0.74	CCBE1	0.68	OPN4	0.63	SLC25A4	0.57	COG5	0.51	CDT1	0.43
LOC115493927	0.92	THBS4	0.81	LOC115490698	0.74	MPP5	0.68	LOC115491180	0.63	IGF2	0.57	GRB2	0.51	LOC115496033	0.43
TLL2	0.91	REL	0.81	LOC115493635	0.73	SLC16A3	0.68	TMEM201	0.63	SLIT3	-0.57	LOC115493910	0.50	INSYN2B	0.43
LOC100229435	0.90	LOC115493865	0.81	ADSS1	0.73	LOC116807734	0.68	TTC28	0.63	LOC100220690	0.57	LOC100223718	0.50	LOC100222957	0.43
LOC115493883	0.90	ZBTB42	0.81	LOC100219000	0.73	SLC13A1	0.68	LOC116808070	0.63	LOC115497877	0.57	SLC25A48	0.50	LOC116809294	0.42
GHR	0.90	BARX1	0.81	ROBO2	0.73	VAV2	0.68	CDCP2	0.63	GFRA1	0.57	ACSF2	0.50	LOC116806919	0.41
SNX10	0.89	LOC100232098	0.80	VSTM5	0.73	SNTG1	0.68	ONECUT1	0.63	TNKS	-0.56	VIT2	0.50	LOC116807261	0.41
LOC115494463	0.89	FAM240A	0.80	LOC115493971	0.72	LOC115491133	0.68	LOC100220758	0.63	CH25H	0.56	WT1	0.49	LOC100220760	0.41
SEMA5A	0.89	LOC115491108	0.80	SPOPL	0.72	TMSB15B	0.67	IL1R1	0.63	WWC2	0.56	LOC115495752	0.49	LOC116807756	0.40
CENPE	0.89	CABP1	0.80	LOC116807829	0.72	AR	0.67	LOC116808082	0.63	LOC116809281	0.56	GABRB1	0.49	TRNAC-GCA_8	0.40
UTS2B	0.89	BCAP29	0.80	DNAH10	0.72	LOC116808110	0.67	YEATS4	0.62	BBS5	0.56	SEPTIN3	0.49	TGIF1	0.40
BMP2	0.89	LOC101234217	0.80	HELT	0.71	POU1F1	0.66	HPRT1	0.62	MCM10	0.56	FLNC	0.49	SHMT1	0.40
LOC115494504	0.88	HERC4	0.79	NME7	0.71	LGMN	0.66	LOC101233072	0.62	TSC22D1	0.56	LOC115494122	0.48	MYPN	0.40
SDC1	0.87	LOC115493758	0.79	PDE9A	0.71	RSPO4	0.66	DIO3	0.62	LOC116808493	0.56	SCGN	0.48	SFRP4	0.40
LRRN4	0.87	FAM185A	0.79	FBXL13	0.71	LOC100221042	0.66	LOC116807890	0.62	AZIN1	0.55	LOC115495359	0.48	LOC116808124	0.39
LOC100232561	0.87	HTR1A	0.78	LOC116808880	0.71	LOC100222749	0.66	LOC115491135	0.62	MARK1	-0.55	LOC115495608	0.48	MIR2984	0.39
LRRC2	0.86	GBX2	0.78	ACSL1	0.71	TACC2	0.66	LOC115496139	0.62	TOM1	0.55	CDYL	0.47	LOC115497917	0.39
LOC115496038	0.86	SIX2	0.78	PTX3	0.71	LOC115495472	0.66	TNFRSF13C	0.62	GMDS	0.55	LOC100222730	0.47	HOXA11	0.38
PTGER4	0.86	PDE1A	0.78	LOC100228395	0.71	MUC2	0.65	ATP11B	0.62	ACKR3	0.55	LOC100220430	0.47	LOC115493492	0.38
MAML2	0.85	LOC116809181	0.78	LOC100222099	0.70	LOC115496281	0.65	LOC116806804	0.61	FAM110A	0.55	TPPP3	0.47	LOC116808892	0.38
LOC100221179	0.85	RGS7BP	0.77	LOC115495208	0.70	LOC105758841	0.65	LOC115494393	0.61	LOC116807701	0.54	LOC100218771	0.47	MMRN2	0.37
LOC115494942	0.85	SLC15A5	0.77	TMC5	0.70	CHRNA5	0.65	ASB15	0.61	LOC116808425	0.54	LOC115496073	0.46	LOC100223936	0.37
LOC116808254	0.85	RRAS2	0.77	LOC116808614	0.70	DSG2	0.65	FOXO3	0.61	NIF3L1	0.54	LOC100218321	0.46	ZNF367	0.37
PPP1R27	0.85	CBX6	0.77	LOC115493739	0.70	NEK10	0.65	LOC100189947	0.61	CHRNB2	0.54	LOC100227844	0.46	LOC115491274	0.37
PHETA1	0.85	LOC100227698	0.77	LOC100232469	0.70	ST6GAL2	0.65	LOC116808684	0.61	PITX3	0.54	NME2	0.46	ACOT11	0.36
MYLK4	0.85	MAN1A2	0.76	GLI2	0.70	DAPL1	0.65	EIF4E3	0.61	MFAP4	0.54	LOC115491148	0.46	BHLHA15	0.36
FGL2	0.85	TAB3	0.76	SLC5A8	0.70	GPHN	0.65	SHISA8	0.61	LOC116808247	0.54	LOC115497656	0.46	LOC115493789	0.35
FOSL2	0.85	SLC7A2	0.76	LOC115490696	0.70	COL19A1	0.64	LOC101233680	0.60	CD83	0.53	PKHD1	0.46	LOC115498168	0.35
FAM102B	0.85	KERA	0.76	LOC100221052	0.70	DMXL2	0.64	TRAPPC9	0.60	CDADC1	0.53	MUC6	0.46	LOC116808559	0.35
ADGRG2	0.85	CCNA1	0.76	LOC116806729	0.70	LOC115494903	0.64	LOC115495769	0.60	PROB1	0.53	LOC115494032	0.46	LOC116808141	0.35
FST	0.84	LMTK2	0.76	UTP25	0.70	CTCFL	0.64	LOC116806809	0.59	LOC100221733	0.53	LOC116808435	0.46	LOC115494754	0.34
LOC115494282	0.84	WDCP	0.75	LOC100219379	0.70	LOC115496575	0.64	LOC115497201	0.59	LOC115492991	0.53	HAO2	0.45	LOC116809267	0.34
MCU	0.84	LOC115493751	0.75	SLC35C2	0.70	LOC116808083	0.64	EYA4	0.59	PKP4	0.53	LOC100223745	0.45	VWA3B	0.34
TRIM67	0.84	LOC100223270	0.75	FFAR4	0.69	ARHGAP24	0.64	LOC100225408	0.59	FOXL3	0.52	LOC115492560	0.45	MIR204-2	0.33
LOC115494992	0.84	SPTLC3	0.75	LOC100223369	0.69	DIPK1A	0.64	LOC100222387	0.59	ASB10	0.52	BRCA1	0.45	GLDN	0.33
LY75	0.83	LOC115495199	0.75	CPA6	0.69	HMMR	0.64	LOC115495471	0.59	LOC115490948	0.52	OPN1SW	0.44	LOC115496310	0.33
FGF18	0.83	LOC116809035	0.75	LOC116809117	0.69	LOC100218870	0.64	LOC116807851	0.58	LOC115496648	0.52	MAP2	-0.44	HPD	0.32
NPTX2	0.82	DMRT1	0.75	CCDC82	0.69	LOC116808795	0.64	LOC100230156	0.58	PGM3	0.52	FOSB	0.44	LOC115491248	0.32
PFKP	0.82	FAM3C	0.74	CCDC146	0.69	TRNAU1AP	0.64	GRIP1	0.58	LOC116807831	0.52	DAPP1	0.44	LOC101233536	0.32
CHRNA3	0.82	LOC115493824	0.74	GPR155	0.69	PDE8B	0.64	LOC115493350	0.58	ITGA9	0.52	TTK	0.44	PLA2G4E	0.31
SCUBE1	0.82	LPAR3	0.74	LOC100227652	0.68	LOC115493570	0.64	HTR1F	0.58	LOC115494131	0.52	LOC105758677	0.44	LOC115492139	0.31
EDIL3	0.81	GPR21	0.74	NFKBIZ	0.68	HEY1	0.64	MUSTN1	0.58	POT1	0.51	TK1	0.43	PRDM13	0.31

568 **Table S1- ModuleG constituent genes.** Lists all genes assigned to module G by and their
569 continuous membership in module G (Pearson r to MEG from moduleG)