Sex chromosome gene expression associated with vocal learning following hormonal manipulation in female zebra finches

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

1: Laboratory 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. Absence of song in females is associated with atrophy and apparent repression of several vocal learning brain regions during development. However, atrophy can be prevented and vocal learning retained in females when given early pharmacological estrogen treatment. To screen for drivers, we performed an unbiased transcriptomic analysis of song learning nuclei and surrounding regions from either sex, treated with vehicle or estrogen until 30 days old when divergence between the sexes becomes anatomically apparent. Analyses of transcriptomes by RNA sequencing identified song nuclei expressed gene modules associated with sex and estrogen manipulation. Female HVC and Area X gene modules were specialized by supplemental estrogen, exhibiting a subset of the transcriptomic specializations observed in males. Female RA and LMAN specialized modules were less dependent on estrogen. The estrogen-induced gene modules in females were enriched for anatomical development functions and strongly correlated to the expression of several Z sex chromosome genes. We present a hypothesis linking loss and de-repression of vocal learning in females, estrogen, and sex chromosomes gene expression.

Introduction

Vocal learning is the ability to imitate sounds, and is a necessary and specialized component for spoken language and song. Vocal learning is found in seven non-human clades, four mammalian and three avian, each having independently evolved the trait[1]. 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 castanotis*). Despite ~300 million years of separation from their common ancestor[2,3], there is remarkable evolutionary convergence between songbird and human vocal learning in terms of behavioral progression, developmental effects of deafening, anatomical connectivity of vocal-motor learning circuits, sites of accelerated evolution within the genome, and specific genes with specialized up- or down-regulated expression in song and speech circuits relative to the surrounding motor control circuits[1,4–10]. Unlike in humans, however, vocal learning is strongly sexually dimorphic in zebra finches and many other vocal learning species[11]. Male zebra finches learn to produce a species appropriate song by imitating mature male conspecifics during juvenile development, while females are limited to producing innate calls[12].

The songbird vocal motor learning circuit contains four major interconnected telencephalic song control nuclei; HVC (proper name) in the dorsal nidopallium (DN); the lateral magnocellular nucleus of the anterior nidopallium (LMAN) 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**)[6]. During juvenile development in zebra finch females, HVC and RA atrophy, HVC fails to form synapses in RA, and Area X never appears[11,13–20]. 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[21–24]. This “masculinized” song system allows estrogen supplemented females to imitate vocalizations, though not with the same accuracy as males[21,23,25–27]. Interestingly, lesioning female HVC prevents estrogen dependent anatomical “masculinization” of its postsynaptic targets of RA and Area X[28].

The genetic basis of this estrogen-sensitive and sexually dimorphic zebra finch vocal learning remains largely unknown beyond the downstream recruitment of the androgen receptor (AR)[29]. However, the examination of a rare gynandromorphic zebra finch with lateralized sex chromosome composition indicates that genetically male HVC and Area X are larger than their female analogs independent of gonadal hormone production, implicating sex chromosome gene expression within the song system[30]. Unlike the mammalian X and Y, birds have a Z and W sex determination system where females are hemizygous (ZW) and males homozygous (ZZ)[31]. The relevant transcriptional machinery appears to be set up by post-hatch day 30 (PHD30), after which estrogen fails to masculinize female song nuclei or behavior and the male song system enlarges while the female song system atrophies[13,15,21,32]. Taken together with the hypothesis that vocal learning in females was lost multiple independent times amongst songbirds[33], the extant findings suggest genetic drivers of vocal learning loss associated with estrogen, sex chromosomes, and song nuclei gene expression specializations. To screen for potential genetic drivers and locate their action within the song system, we performed an unbiased analysis of transcriptomes from song nuclei and surrounding motor control regions in zebra finches of either sex chronically treated with 17-β-estradiol (E2) or vehicle from hatch until sacrifice at PHD30 (**Fig. 1b**). We used a new zebra finch genome assembly and annotation produced by the Vertebrate Genomes Project containing both the Z and W chromosomes[34].

Results

Identification of gene expression modules in song nuclei

We first sought to characterize differences in song nuclei gene expression specializations relative to surrounds in juvenile males and females with and without E2 treatment. We used RNA-Seq data from a previous study by our lab, on the effects of E2 manipulation on the song system in both males and females[25]. This E2-treatment program masculinized females sufficiently for them to produce song as adults in another cohort of birds[25]. Birds were sacrificed as juveniles at post-hatch day 30 (PHD30), after a one hour period of silence to limit activity-dependent gene expression in song nuclei and surrounding motor regions; this developmental age was chosen because it is around the time when both males and E2-treated females start to sing (e.g. subsong) and all four song nuclei are sufficiently developed to be visually apparent in histological sections[25]. The four major song nuclei (HVC, LMAN, RA, and Area X) and their adjacent surrounding motor regions (DN, AN, LAI, and Str respectively; **Fig. 1a**) were dissected using laser capture. The E2-treated females in this study had similarly sized song system nuclei as males, indicating that E2 treatment prevented atrophy. This past study mapped the RNA-Seq reads to an older genome assembly lacking the W sex chromosome[25].

With this data, we first remapped RNA-Seq reads (from n = 3 birds per group) to a new zebra finch assembly (GCF_008822105.2) with both the Z and W chromosomes. In our quality control analyses, hierarchical clustering of sample expression vectors revealed two technical outliers, one HVC from a vehicle-treated male HVC and one RA from a E2-treated female, which we excluded from further analysis (**Fig. S1**). Similar to previous PCA and hierarchical clustering[25], the remapped RNA-Seq data expression levels with outliers removed still resulted in separation of song nucleus and surround and some E2-treated female samples, without sufficient resolution at finer group levels (**Fig. S1**). To address our questions at finer resolution and in an unbiased way, we performed weighted gene correlation network analysis (WGCNA), where we decomposed the transcriptome into actively expressed gene modules using hierarchical clustering of the gene adjacency matrix. This matrix describes the inferred structure of gene networks within our data and was calculated by soft thresholding the matrix of gene-to-gene expression correlations across samples (**Fig. S2**)[35]. After initial hierarchical cluster-based module assignment, genes were iteratively reassigned to the module whose aggregate expression they best correlated with until no additional genes met the WGCNA reassignment threshold (**Fig. S3** and **Methods** for parameterization). This optional reassignment was performed to ensure that genes are matched to the aggregate measure that best represents their expression.

Of the ~21,000 annotated zebra finch genes, 13,220 were well expressed in the finch telencephalic brain regions sampled, a comparable number of genes to what we have seen expressed in adult zebra finch telencephalon[9,36]. These 13,220 were assigned to 14 co-expression modules (**Fig. S4a**). The median observation of unassigned transcripts was 22-fold lower than the median observation of module assigned transcripts (4.86 vs. 0.22 FPKM). Two modules clearly marked single samples from two different birds (**Fig. S4a,b - arrows**), indicative of technical overfitting; these two modules (not birds) were excluded from further analysis, resulting in a more visibly diverse pattern within and across modules (**Fig. S4c,d**). The remaining 12 modules contained 12,444 genes in total were lettered in descending order of size A through L, containing from 4,890 to 127 constituent genes each, and were dynamically expressed across brain regions and treatment groups (**Fig. 1c**)

Lack of male HVC and Area X specializations in juvenile females

To understand song system transcriptional specialization in the context of the gene modules, we calculated the module eigengene (MEG) expression values for each of the 12 modules (**Fig. 1d**). MEGs are the first principle component of variance of all genes in a module and are the aggregate measure for each module's expression across samples. We then tested for statistically significant correlations between MEG expression for each module and the song nuclei specializations relative to their respective surrounds. Each song nucleus had unique and overlapping specialization of genes in 2 to 4 modules of the 12 (**Fig. 2a)**. The LMAN specialization relative to the surrounding AN correlated with the expression of gene modules B, F and I; HVC specialization relative to DN correlated to the expression of gene modules B, F, and G; RA specialization relative to LAI consisted of gene modules C and L; and Area X specialization relative to the surrounding Str consisted of gene modules C, F, G, and I. Area X was the only song nucleus that did not have a specialized module unique to it relative to other song nuclei (**Fig. 2a**).

In non-vocal learning juvenile females, interestingly LMAN was specialized relative to the AN by the same gene modules as in males (B, F, and I) as well as an additional module G (**Fig. 2b**); RA was specialized by module A as in males, but not module L and by additional modules A and G. In contrast, neither juvenile female HVC nor Area X exhibited significant gene module expression specializations relative to their surrounds. We interpret these findings to indicate that: LMAN and RA atrophy later in

juvenile female development; that module A and G genes could be associated with the start of this atrophy; HVC and Area X are likely the first to atrophy or not develop; and lack of any gene module specialization in them at this age could mean that they would be more sensitive to estrogen prevention of vocal learning loss.

E2-responsive gene modules in song nuclei

To assess the effects of chronic exogenous estrogen on the developing song system, we first performed a control analysis of modules in the E2-treated juvenile males. In the E2-treated juvenile males, the song nuclei specialized modules overlap to those seen in vehicle-treated males (**Fig. 2a vs 2c**); differences were that module I in LMAN and Area X and module L in RA were no longer present; and module G newly appeared in HVC, module A appeared in RA as seen in females, and module E also appeared in RA. These differences in gene module specializations in the male song system are consistent with the E2-treatment regimen used causing a slight decrease in vocal learning accuracy in males[25].

In contrast, E2-treated females more closely mirrored the gene module specializations seen in the vehicle treated male song system (**Fig. 2a vs 2d**). Specifically, module G was not associated to RA while module C became more strongly correlated; module G expression was strongly specialized in HVC; and module C appeared specialized in Area X. Despite these changes to be more like males, in E2-treated females, module A in RA maintained its specialization as seen in vehicle-treated females (**Fig. 2b vs 2d**), and module E appeared in RA as in E2-treated males (**Fig. 2c vs 2d**); a weak correlation of a module I in HVC and module K in RA appeared (**Fig. 2d**). That is, the transcriptional response of female song nuclei to E2-treatment appeared to be far more dramatic than in males, consistent with estrogen inhibition in males having only a limited impact on song nuclei and vocal learning[25].

We performed an additional test for E2-induced changes to gene module expression in females by comparing E2-treated song nuclei in females to the combination of E2-treated surrounds from the same animals, vehicle-treated female surrounds, and vehicle-treated female song nuclei (**Fig 2e**). This analysis represents a comparison between samples from vocal learning capable female samples and non-vocal learning female samples at each node. It revealed 1 to 3 modules in each song nucleus of E2-treated females that were a subset of some of the strongest correlated modules found in control males (**Fig. 2a vs 2e**). These were: module F in LMAN; module G in HVC; module C in RA; and modules C, F and G in Area X. That is, these are four core modules for song nuclei that are the most sensitive to E2-induction or enhancement in females and most associated with the presence of vocal learning in females.

Gene modules for telencephalic sex differences

We next tested whether some modules could be explained by sex differences in the brain. We compared vehicle treated male and female song nuclei and surrounds separately, and found that module E strongly correlated with higher expression in all male brain regions relative to females (**Fig. 2f,g**). Among the song nuclei, module G eigengene expression was significantly higher in male HVC and Area X; conversely, module C expression was higher in female LMAN and RA (**Fig. 2f**). Among the surrounds, module H eigengene expression (not significant in any other comparison) was significantly higher in female DN and module C was significantly higher in female AN and Str (**Fig. 2g**). These

findings indicate the existence of a sex specific, telencephalon-wide shift in the transcriptome represented in these data by module E. This broad sexual dimorphism is further refined by regionally patterned sex differences like those observed for module G in HVC.

Gene modules with region specific expression

We next quantified the magnitude of module expression differences and observed several gene modules with clear region-specific expression patterns (**Fig. 2h-k**). Module A, which appeared specialized in RA relative to LAI in several comparisons, was highly expressed in all Area X and Str samples, regardless of treatment or sex (**Fig. 2h**). Interestingly, *FOXP2*, a gene critical for spoken-language and vocal learning in humans[37] and songbirds[38], was a potent member of module A, correlating with the eigengene at $r^2 = 0.92$ across all samples. Module C, which was specialized to several song nuclei including RA, HVC, and Area X relative to their surrounds and sexually dimorphic in LMAN and Str, was most highly expressed in arcopallial samples, especially RA, with some increase in HVC (**Fig. 2i**). Module F, which was specialized to LMAN relative to its surround in all such comparisons, was highly expressed exclusively in LMAN regardless of sex and E2-treatment (**Fig. 2j**). Module G, which was specialized to HVC in a sex and E2 dependent manner, was only highly expressed in HVC samples, with males having higher expression than females, and females further split by treatment (**Fig. 2k**).

Functional enrichment of specialized modules

We sought to understand the cumulative biological function of genes among the modules. To do this, 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 each module's constituent genes. Of the 12,444 module assigned genes, 7,909 (63%) had 1:1 human orthologs annotated in Ensembl. We found GO terms significantly enriched in module G (**Table 1**), 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 specialized in male and E2-treated female HVC and Area X potentially enact a proliferative developmental program. Other significantly enriched terms were "extracellular matrix structural component", "external side of the plasma membrane", and "extracellular space", indicating that this module may also act to restructure the extracellular matrix, perhaps to accommodate new cells. Module E, which was differential between the sexes for both song nuclei and surrounds, had 6 significantly enriched terms, of which 3 pertained to DNA damage repair: "nucleolus", "transcription, DNA templated", and "U2-type precatalytic spliceosome" (**Table 2**). These results in module E indicate that there is a telencephalon-wide sexually dimorphic gene expression program in finches that likely acts within the nuclear environment. A full table of GO enrichments by module can be found in supplemental data (supplemental_sig_go_enrich.csv).

Gene modules enriched for human speech associated genes

We next asked if any of the modules were enriched for genes previously determined to be convergently specialized in songbird song nuclei and human speech brain regions[7,9]. We performed this analysis by treating the different sets of convergently specialized genes as gene ontology terms. These gene lists

included the transcriptional convergence between RA and human dorsal laryngeal motor cortex (dLMC), HVC and dLMC, LMAN and dLMC, Area X and the anterior caudate, and Area X and the anterior putamen (Gedman et al. Fig3B,D[9]) . We found that each of these convergent gene lists was enriched for the modules specialized to or highly expressed in the finch brain region used to generate the list (**Fig3a**). We found that module B, specialized to both LMAN and HVC (**Fig 2a**), was enriched for the convergently specialized genes in human dorsal laryngeal motor cortex (dLMC; **Fig 3a**)[9]. Module C, specialized to RA (**Fig 2a,i**), was enriched for the genes convergently specialized in dLMC and RA (Fig 3a) known to match the lower layers of the cortex[9]. Module A, highly expressed in Area X and Str (**Fig 2h**), was enriched for genes convergently specialized in Area X and human anterior striatum (both caudate and putamen; Fig.3a). ; Module I, specialized to Area X (**Fig 2a**), was also even more strongly enriched for the same convergences as module A (**Fig 3a**). These findings indicate that the genes previously identified as convergently regulated between songbird song brain regions and human speech brain regions are components in the larger specialized gene networks identified here using WGCNA.

Gene modules enriched for specific chromosomes

We next determined whether any modules were enriched in genes from specific chromosomes. We performed a bootstrapped analysis, randomizing the mapping between genes and modules 50,000 times to approximate null distributions. P values for each chromosome-module pairing were then FDR corrected. To do this as accurately as possible, we performed this analysis using revised chromosomal assignments and structure from the most recent zebra finch genome assembly, bTaeGut_1.4.pri, which better resolves the microchromosomes[39] (see **Fig. S6** for the less resolved bTaeGut_W.pat.v2 assembly). Surprisingly, we found that each of the 12 modules were enriched for genes from at least one chromosome (**Fig. 3c**). The most striking enrichment was of module E genes on the Z and W sex chromosomes, with nearly all expressed W genes and ~⅔ of expressed Z genes being members of module E. This is consistent with module E exhibiting higher expression in the male telencephalon regardless of treatment or region (**Figs. 2d-e**). For the autosomes, we observed significant enrichments of two HVC specialized gene modules on chromosome 1A; module B (specialized in male HVC and human dLMC); and module G (specialized in male HVC and E2-treated female HVC and sexually dimorphic in vehicle-treated HVC). These results are particularly interesting given that zebra finch chromosomes 1 and 1A are believed to be the result of a songbird-specific fragmentation of the ancestral chromosome 1 found in chickens[40,41].

Module A, the largest module which was highly expressed in Area X and adjacent striatum, was enriched across 5 macro-chromosomes (chr1, 2, 3, 4, and 10). Module F, which was strongly expressed in LMAN, was enriched on chromosomes 2 and 7. Modules C (most strongly expressed in the arcopallium and part of RA and Area X specializations) and I (a component of LMAN specialization) was enriched on chromosome 6 on microchromosome 35. Module D, a component of the LMAN specialization in E2-treated females, was enriched across 8 small microchromosomes (chr22, 23, 25, 27, 28, 29, 30, 31). 4 of the 5 smallest modules in gene counts were enriched in the microchromosomes: Module H, which was sexually dimorphic only in DN, was enriched on microchromosome 28; module K was significantly enriched on microchromosome 37 and in the “other” category, which includes all remaining unnamed DNA scaffolds in the assembly such as further microchromosomes; module J was also enriched in “other”.

Sex- and micro-chromosome gene expression across the telencephalon

The telencephalon-wide, sexually dimorphic expression of the sex chromosome dominated module E is one of the strongest features in the present dataset; Module E eigengene is highly expressed in all male samples and depleted in all female samples regardless of brain region or treatment (**Fig. 4a vs 4c-d**). To better understand the relationship between the sex chromosomes and module E, we examined the distribution of membership in module E with the sex chromosomes separated. WGCNA allows us to consider gene membership in a module as a continuous variable, rather than a binary variable, by correlating each gene's expression profile to the module eigengene. Doing this for module E, we found that Z transcripts were positively correlated to the module eigengene while W transcripts were anticorrelated (**Fig. 4e**). This is consistent with Z chromosome transcripts being generally depleted in females relative to males, while W chromosome transcripts were only expressed in female brains. This reduction in Z chromosome transcript abundance is consistent with the finding that diploid Z chromosomes in male birds do not have one copy inactivated to compensate for gene dosage, unlike the X inactivation in female mammals[30,42].

Given the sexually dimorphic expression of module E across brain regions and the enrichment of the sex chromosomes within that module, we separately compared the expression of sex chromosome genes in the non-vocal motor surrounds of vehicle-treated control males and females regardless of WGCNA assignment to better understand sex chromosome expression without the influence of vocal learning specialization or E2 response. To do this, we correlated expression of each sex chromosome gene to the sexual dimorphism, such that female-enriched transcripts were positively correlated and male-enriched transcripts were anticorrelated (**Fig. 4f,g**). In these gene sets, we identified between 73 and 82 significantly expressed W chromosome genes and between 433 and 527 significantly depleted Z chromosome genes for each of the surround regions in males relative females. Examining the union of these gene sets revealed that a total of 95 W and 694 Z chromosome genes were differentially expressed between sexes in at least one brain region, 62% and 65% of annotated sex chromosome genes respectively. Conversely, the intersection of these regional gene sets contained 58 significantly expressed W chromosome genes and 306 significantly depleted Z chromosome genes in all non-vocal regions in females, representing 38% and 29% of annotated sex chromosome genes, respectively (**Fig. 4h-i**, **Table S2**). This indicates that there is a telencephalon wide set of sex enriched and depleted sex chromosome genes in addition to regionally patterned sex chromosome gene expression.

We also observed strong enrichments of several of the smaller gene modules, J and K, on microchromosomes. Examining the MEG expression for these modules, we found that both were expressed brainwide in non-overlapping sets of animals. Module J was highly expressed in all samples taken from animal-b, an E2-treated female (**Fig. 4b**); similarly, module K was most highly expressed across all samples from animal-g, an E2-treated male, and exhibited intermediate expression in all samples from animal-f, a vehicle-treated female (**Fig. 4c**). Though not enriched on microchromosomes, Module L exhibited similar animal-specific high expression in animal-a, an E2 treated female (**Fig. 4d**). These animal and chromosome specific shifts in the transcriptomes could represent the systemic effects of allelic chromosomal structural variation which WGCNA was able to capture due to our broad sampling of regions across the telencephalon.

Modeling vocal learning and sex chromosome module interactions

As it is unlikely that the gene modules act independently of each other, we sought putative interacting genes between two modules of interest, module G specialized in HVC and Area X and module E dominated by sex chromosome genes. To do this, we again replaced binary in-or-out module membership with continuous module membership by correlating each genes' expression to the relevant module eigengene[35]. This allowed us to quantify the extent to which any gene was associated with any module, regardless of initial assignment[33]. Looking across all assigned genes for our modules of interest, we identified two outlier genes, *PDE8B* and *HABP4*, which were the most E associated gene assigned to module G and the most G associated gene assigned to module E, respectively (**Fig. 4j**). Both genes were significantly upregulated in male HVC relative to its surround, but not in female HVC of either treatment. *PDE8B* catalyzes the hydrolysis of the second messenger cAMP and mutations to the gene cause an autosomal dominant form of striatal degeneration in humans[43]. *HABP4* is an RNA binding protein, known to repress the expression and subsequent DNA binding of *MEF2C*[44], a Z chromosome transcription factor which has undergone accelerated evolution in songbirds[10] and whose repression by *FOXP2* is critical for cortico-striatal circuit formation in mice related to vocal behaviors[45]. Both *PDE8B* and *HABP4* are found on the Z chromosome and *HABP4* was one of the 694 Z transcripts significantly depleted across the telencephalon in vehicle-treated females relative to males (**Table S2**). These findings indicate that Z chromosome genes may be subject to multiple gene regulatory programs; both the telencephalon wide depletion driven by reduced copy number and specialized upregulation in HVC for vocal learning behavior.

Candidate gene drivers of HVC specialization in E2-treated females

To reduce the 344 genes in module G to the putative drivers of masculine HVC development, we next examined the relationship of continuous membership in module G (correlation to MEG-G) to gene expression specialization in HVC at the level of single genes. We did this by correlating gene expression to the specialization of HVC in males (vehicle- and E2-treated) or E2-treated females for each of the following four vocal learning comparisons: 1) male HVC specialization relative to the surround; 2) E2-treated female HVC specialization relative to the surround; 3) HVC sexual dimorphism in vehicle-treated controls; and 4) E2-treated female HVC specialization relative to vehicle-treated female HVC. For all four comparisons, we found the higher the correlation of module G genes to the MEG, the higher the correlation with the vocal learning specialization (**Fig. 5a-d**). This means that their expression was higher in male or E2-treated female HVC relative to the appropriate non-vocal learning controls across all comparisons. We identified genes of interest as being strongly correlated ($r^2 \geq 0.5$) to both the module G eigengene and vocal learning specialization (**Fig. 5e-h**, higher magnification view of colored boxes in **Fig. 5a-d**). All genes of interest exhibited a positive correlation to vocal learning specialization across all four comparisons (**Fig. 5a-d**). We next generated a core gene list for module G involvement in E2- and sex-dependent HVC development by intersecting these four vocal learning specialized gene sets (**Fig. 5i**). We found a core set of 14 genes that strongly marked vocal learning capable HVC in all comparisons and strongly correlated with the aggregate of module G expression (**Table 3, Fig. S6**).

We examined the chromosomes of origin for these 14 core genes and found that three (*GHR*, *RGS7BP*, and *THBS4*) were on the Z sex chromosome (**Fig. 5j**). This was a >3-fold significant enrichment over chance of Z chromosome transcripts ($p = 0.009$, 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 three Z chromosome module G genes in HVC not only exhibited >50% reduced expression in control female HVC relative to males, but also exhibited upregulated expression in male HVC relative to the

surrounding DN regardless of treatment and upregulation in E2-treated female HVC relative to either the surround or vehicle-treated female HVC (**Fig. S7**). In the context of our cross-region sex chromosome analysis (**Fig. 3c-f**), only *RGS7BP* was sexually dimorphic outside of the song system, being depleted in all female surround regions. Neither *GHR* (growth hormone receptor) nor *THBS4* were significantly depleted in any surround region comparison. Taken together, these results indicate that the 3 core genes in vocal learning capable HVC on the Z chromosome are subject to additional E2 sensitive transcriptional regulation in HVC, separate from the Z chromosome transcript depletion seen throughout the female brain.

Of the 14 core genes total, several have been previously studied in the brains of other species which may inform their role in vocal learning systems. *THBS4* encodes a secreted extracellular-matrix glycoprotein necessary for appropriate neuronal migration in the mouse[46] and is elevated 6-fold in the human cortex compared with non-vocal learning primates[47]. Human *EDA2R* was recently identified as a top correlate of cognitive performance and brain size[48]; it was also found in a human GWAS study to correlate with circulating estrogen and testosterone levels[49]. Rare mutations in human *PHETA1* lead to Lowe oculocerebrorenal syndrome, that includes pathophysiology in seizures, mental retardation, and structural brain abnormalities[50,51]. *SIX2* is a homeobox domain containing transcription factor that governs early brain and craniofacial development and provides neuroprotection from dopamine injury[52,53]. *GHR* encodes a transmembrane receptor whose activation controls cell division[54]. The gene which encodes GHR's ligand, growth hormone, is interestingly duplicated and undergoing accelerated evolution in the genomes of songbirds, including the zebra finch[55]. Based on these findings, we consider *GHR* as the most likely candidate gene related to the E2 sensitive atrophy of HVC in females. Based on the findings in these gene sets, we hypothesize that without excess estrogen in females, HVC expansion is prevented by not specializing the growth and neuronal migration promoting genes in module G to the HVC lineage by late development. This is potentially enacted by depleting necessary gene products from the Z sex chromosome, such as GHR, which are already present in only one copy.

Discussion

The present study seeks to further our understanding of sexually dimorphic vocal learning in zebra finches by comparing the gene expression specializations in the song system between male and female birds at PHD30, during the critical period of atrophy for the female vocal circuit. The birds were given either a vehicle or E2 from hatching, which induces rudimentary vocal learning behavior in females where it would otherwise be absent. HVC appeared unspecialized at the level of gene module expression in control females, while in E2-treated females HVC exhibited a subset of the observed male HVC gene expression specialization. Similarly, Area X appeared unspecialized in vehicle-treated females, but exhibited a subset of the male Area X specialization in E2-treated females, consistent with the known absence of Area X in vehicle-treated females and presence in E2-treated females. This contrasts with RA and LMAN which were similarly specialized in males and females in the absence of E2-treatment. Given that lesions of HVC prevent the emergence of Area X in E2-treated females[28], these results support a model of zebra finch development where transcriptomic masculinization of female HVC by E2 is a critical event which facilitates the emergence of female Area X and ultimately endows these females their rudimentary song. When examining the module G HVC specialization induced by E2-treatment in female HVC, we surprisingly found that the most specialized genes were disproportionately from the Z chromosome.

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? One

possibility is that this process begins with increased estrogen receptor (ER) activity within developing HVC cells after being provided surplus activating ligand[56], followed by altered transcription of ER targets in the genome. Module G contained the androgen receptor (**Table S1**), which is believed to be a major downstream effector of E2 response in female zebra finches[29]. This initial transcriptional loading of the system would then have been processed by gene regulatory networks within each cell, spreading in effect through the transcriptome. It is possible that differential module G expression arose purely from traditional gene regulatory networks, where transcription factors form complex, elaborate feedback networks with themselves and the genes they regulate. However, this framework fails to explain why the core 14 vocal learning correlated genes from module G in HVC were enriched for transcripts from a single chromosome: the Z sex chromosome with halved copy number in females.

We hypothesize that the Z chromosome genes identified here are co-regulated, necessary components of a proliferative transcriptional program downstream of ER, represented by module G. 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 maintained through persistent *GHR* signaling. We propose that these Z chromosome transcripts in module G are depleted in females by lower haplotype dosage during development and thus fail to specialize female HVC. Due to lower abundance of gene products from module G, the female HVC could be less proliferative during juvenile development and fail to facilitate emergence of its downstream target Area X. Similarly, without a sufficiently developed HVC, RA lacks one of its major inputs and may subsequently atrophy. We further propose that E2 masculinizes female song behavior by increasing the abundance of these module G transcripts in HVC, increasing specialized HVC proliferation, and facilitating emergence of HVC's other major target, Area X (**Fig 6**). This model of sex chromosome influenced song system development is consistent with recent work comparing male and female zebra finch transcriptomes from RA at young juvenile (PHD20) and young adult (PHD50) ages in un-manipulated birds, where they found that sexually dimorphic Z chromosome expression at juvenile ages preceded the dimorphic expression in the autosomes seen in adults[57].

Our results help refine the traditional notion that hormonal signaling organizes the brain to produce sexually dimorphic behaviors independent of neuronal sex chromosome content[58,59]. Instead, our data indicate that sexual dimorphism in zebra finch song learning ability was likely established by the interaction of sex hormone signaling and sex chromosome gene expression within HVC during development. These findings are thematically similar to work in the Four Core Genotypes mouse model[60], where chromosomal and gonadal sex are separable by translocating the sex determining *Sry* gene. In these mice, sex chromosome composition regulates sexually dimorphic brain gene expression, circuit anatomy, and behaviors[60–63], though the sex chromosome genes responsible remain unknown. Additional experiments manipulating the candidate genes implicated here in developing HVC to both mimic and prevent the action of E2 in female zebra finches are needed to test these hypotheses.

In addition to these vocal learning focused results, performing unbiased gene network analysis with samples taken throughout the zebra finch telencephalon in both sexes revealed a surprisingly strong relationship between the modular organization of telencephalic gene expression and the chromosomal structure of the genome. Each of the 12 modules identified by WGCNA were enriched for genes from at least one chromosome. Of these significant enrichments, those from on the micro- and sex-chromosomes stood out as particularly strong. Module E encompassed the majority of W and Z chromosome genes and was sexually dimorphic in its expression throughout the telencephalon. Indeed, we found that roughly ⅓ of sex chromosome transcripts are significantly sexually dimorphic in all non-vocal brain regions, and roughly ⅔ are sexually dimorphic in at least one region. This half-brain-wide and half-regionally-patterned sexual dimorphism observed from the sex chromosomes may act as a substrate for the evolution of sexually dimorphic behaviors generally, with brain-wide shifts

in expression providing a transcriptional background for evolution of sex-specific patterning in additional sex chromosome genes. These regionally patterned sex chromosome genes could then recruit gene expression networks from somatic chromosomes, resulting in the sex specific anatomical patterning of gene expression from throughout the genome. This general model is consistent with local depletion of GHR and other Z chromosome transcripts in the developing HVC lineage, leading to a loss of specialized module G expression from somatic chromosomes. This hypothesis of an interaction between sex chromosomes and autosomes linked to the gain and loss of vocal learning in one sex, can be tested in future studies.

Acknowledgments

Funding for this work was provided by the Howard Hughes Medical Institute (EDJ), NIH-NIDCD R01-DC016224 (HM), and the NSF-GRFP (MHD). We thank Gregory Gedman, Giulio Formenti, Caitlin Gilbert, Lindsey Cantin, César Vargas, Chul Lee, and Jason Manley for conversations during visualization and analysis. We also thank Alipasha Vaziri and Tobias Nöbauer for providing the computing infrastructure used throughout. This analysis is a memorial to Mark Konishi (1933-2020) whose work on this topic influenced us greatly.

Author contributions

MHD performed all analyses, generated all plots, and wrote the paper; HNC gathered all data and edited the paper; HM and EDJ edited the paper.

Materials & Correspondence

Correspondence to Matthew H. Davenport and Erich D. Jarvis

Competing interests

HM has received royalties from Chemcom. HM has received research grants from Givaudan. HM has received consultant fees from Kao.

Materials and Methods

Animal Handling and Sample Preparation:

The 96 samples used in the present analysis are the E2 or vehicle treated subset of a previously published RNAseq dataset[25]. We briefly re-describe our methodology here. All animal procedures were approved by the IACUC of Duke University.

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[25].

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. No sample pooling was performed, each sample originated from a single bird. This is 8 samples worth of RNAseq data per bird for 12 birds, providing 3 samples per sex-treatment-region combination, roughly a terabyte of read data.

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[64] and then mapped to a high-quality Vertebrate Genomes Project (VGP) female zebra finch nuclear genome (bTaeGut2.pat.W.v2, GCF_008822105.2)[34] using STAR (v2.7.1)[65]. Uniquely mapped reads were then tallied at the level of genes using Rsubread::featureCounts (R-3.6.3)and then counts normalized to fragments per kilobase of transcript per million mapped reads[66]. Multi-mapped reads were rejected as they have a higher probability of being associated with technical artifacts of sequencing or genome assembly. Read based quality control was performed with FastQC (Babraham Bioinformatics) with reports prepared by MultiQC (Python-3.5.5)[67]. 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-4.2 unless otherwise specified. Data was wrangled in the tidyverse, and custom visualizations produced with ggplot, ggdendro, VennDiagram, RColorBrewer, and ggpubr[68,69]. Unsigned topological overlaps between genes (gene-to-gene correlations) were calculated in a single block with WGNCA::blockwiseModules with a soft thresholding power (β) of 6 based on scale free topology fit and model connectedness as described in the WGNCA vignette (**Fig. S2)**[35]. Next we determined an appropriate WGCNA parameterization quantitatively and qualitatively by sweeping the module size and tree cut height parameters of WGNCA::recutBlockwiseTrees. We selected our model for analysis by examining the resulting gene assignments and sample-sample distance matrices. We were able to increase or decrease the proportion of genes assigned to WGCNA modules by lowering or raising the minimum module size parameter, respectively. We found that setting the minimum size parameter below 100 genes included more genes, but with models that increasingly

over-fit single samples, producing obvious outliers in the distance matrix (**Fig. S3**). Correspondingly, raising the minimum size parameter beyond 100 included fewer genes, but did not greatly reduce the number of outlier samples. Based on this we selected 100 as the minimum module size, parameterizing to explain as much transcriptomic variance as possible while minimizing the number of technically overfit samples (**Fig. S2-4**).

Module association to vocal learning:

Module eigengenes (MEGs) from each module were correlated against binarized song system membership, vocal learning capability, or sex and the statistical significance of each correlation assessed using WGCNA::corPvalueStudent. This was done in the following sample subsets by node: male samples broken out by treatment; female samples broken out by treatment; all female samples; song system components from either sex treated with vehicle; and surrounding control regions from either sex treated with vehicle. Within each of the 4 sex-treatment combinations we compared the song system components to surrounds at each node. Within all female samples we compared the vocal learning capable E2-treated song system elements to all other female samples from each node. Within the vehicle song systems and vehicle surrounds we compared between male and female finches for each region.

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. Uncorrected p values for the enrichment of human gene ontology terms within the human orthologs of module G were calculated using generally applicable gene set enrichment (GAGE) implemented in gage::gage[70]. To determine if the genes previously shown as convergently differentially expressed in the zebra finch song system and human speech brain regions mapped to specific modules, we treated these five gene lists identically to GO terms and tested for their significant enrichment across the human orthologs of each module also using GAGE.

Analysis of sex chromosome gene expression independent of vocal learning or E2 treatment:

Sex chromosome transcripts, regardless of WGCNA module assignment, were examined in the vehicle-treated non-vocal learning producing surround samples for each node. To find the most consistently expressed and depleted W and Z chromosome genes respectively, we correlated expression of each sex chromosome transcript with sexual dimorphism within each region, such that expressed W genes would be positively correlated and depleted Z chromosome genes would be anticorrelated. We computed correlations and p values using the WGCNA corAndP function; upper-tailed for the W chromosome and lower-tailed for the Z chromosome. Genes significantly expressed or depleted across regions were then intersected to identify consistently regulated transcripts across the telencephalon.

Module enrichment on chromosomes:

To associate modules to chromosomes, we bootstrapped FDR corrected p values for the enrichment of each chromosome-module pairing by randomizing the mapping of genes to modules 50k times and calculating the enrichments observed on each chromosome from each module in each randomization to empirically determine the null distributions. The calculation of bootstraps and p values was performed in Python-3.5.5 and parallelized using joblib's Parallel.

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

We defined genes of interest as the subset of significantly (T test for correlation: $p \leq 0.05$) 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, calculated using WGCNA::corAndP. 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.

Data and Code Availability

All raw data for this experiment is available on the NCBI Sequence Read Archive (accession: PRJNA698257). The count matrix, quality control results, and analysis code is available online (https://github.com/mattisabrat/sex-and-song).

References

1. Jarvis, E. D. Evolution of vocal learning and spoken language. *Science* **366,** 50–54 (2019).
2. Kumar, S. & Hedges, S. B. A molecular timescale for vertebrate evolution. *Nature* **392,** 917–920 (1998).
3. 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).
4. Bolhuis, J. J., Okanoya, K. & Scharff, C. Twitter evolution: converging mechanisms in birdsong and human speech. *Nat. Rev. Neurosci.* **11,** 747–759 (2010).
5. Doupe, A. J. & Kuhl, P. K. Birdsong and human speech: common themes and mechanisms. *Annu. Rev. Neurosci.* **22,** 567–631 (1999).
6. Mooney, R. Neural mechanisms for learned birdsong. *Learn. Mem.* **16,** 655–669 (2009).
7. Pfenning, A. R., Hara, E., Whitney, O., Rivas, M. V., Wang, R., Roulhac, P. L., Howard, J. T., Wirthlin, M., Lovell, P. V., Ganapathy, G., Mouncastle, J., Moseley, M. A., Thompson, J. W., Soderblom, E. J., Iriki, A., Kato, M., Gilbert, M. T. P., Zhang, G., Bakken, T., Bongaarts, A., Bernard, A., Lein, E., Mello, C. V., Hartemink, A. J. & Jarvis, E. D. Convergent transcriptional specializations in the brains of humans and song-learning birds. *Science* **346,** 1256846 (2014).
8. Feenders, G., Liedvogel, M., Rivas, M., Zapka, M., Horita, H., Hara, E., Wada, K., Mouritsen, H. &

Jarvis, E. D. Molecular mapping of movement-associated areas in the avian brain: a motor theory for vocal learning origin. *PLoS One* **3,** e1768 (2008).

9. Gedman, G. Songbird brain organization and its molecular convergence with humans for vocal imitation learning. (2021).
10. 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.*
11. Nottebohm, F. & Arnold, A. P. Sexual dimorphism in vocal control areas of the songbird brain. *Science* **194,** 211–213 (1976).
12. The zebra finch: a synthesis of field and laboratory studies. *Choice* **34,** 34–5095–34–5095 (1997).
13. 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).
14. 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* **521,** 326–341 Preprint at https://doi.org/10.1002/cne.23174 (2013)
15. 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).
16. 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).
17. 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).
18. 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).
19. 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).
20. 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).

21. Gurney, M. E. & Konishi, M. Hormone-induced sexual differentiation of brain and behavior in zebra finches. *Science* **208,** 1380–1383 (1980).
22. Gurney, M. E. Behavioral correlates of sexual differentiation in the zebra finch song system. *Brain Research* **231,** 153–172 Preprint at https://doi.org/10.1016/0006-8993(82)90015-4 (1982)
23. 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).
24. 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).
25. Choe, H. N., Tewari, J., Zhu, K. W., Davenport, M., Matsunami, H. & Jarvis, E. D. Estrogen and sex-dependent loss of the vocal learning system in female zebra finches. *Horm. Behav.* **129,** 104911 (2021).
26. Pohl-Apel, G. The correlation between the degree of brain masculinization and song quality in estradiol treated female zebra finches. *Brain Res.* **336,** 381–383 (1985).
27. Pohl-Apel, G. & Sossinka, R. Hormonal determination of song capacity in females of the zebra finch: Critical phase of Treatment. *Z. Tierpsychol.* **64,** 330–336 (1984).
28. 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).
29. Nordeen, K. W., Nordeen, E. J. & Arnold, A. P. Estrogen establishes sex differences in androgen accumulation in zebra finch brain. *J. Neurosci.* **6,** 734–738 (1986).
30. Agate, R. J., Grisham, W., Wade, J., Mann, S., Wingfield, J., Schanen, C., Palotie, A. & Arnold, A. P. Neural, not gonadal, origin of brain sex differences in a gynandromorphic finch. *Proc. Natl. Acad. Sci. U. S. A.* **100,** 4873–4878 (2003).
31. Gianaroli, L., Magli, M. C. & Ferraretti, A. P. in *Brenner's Encyclopedia of Genetics (Second Edition)* (eds. Maloy, S. & Hughes, K.) 397–400 (Academic Press, 2013).
32. 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).
33. Odom, K. J., Hall, M. L., Riebel, K., Omland, K. E. & Langmore, N. E. Female song is widespread and ancestral in songbirds. *Nat. Commun.* **5,** 3379 (2014).

34. Rhie, A., McCarthy, S. A., Fedrigo, O., Damas, J., Formenti, G., Koren, S., Uliano-Silva, M., Chow, W., Fungtammasan, A., Kim, J., Lee, C., Ko, B. J., Chaisson, M., Gedman, G. L., Cantin, L. J., Thibaud-Nissen, F., Haggerty, L., Bista, I., Smith, M., Haase, B., Mountcastle, J., Winkler, S., Paez, S., Howard, J., Vernes, S. C., Lama, T. M., Grutzner, F., Warren, W. C., Balakrishnan, C. N., Burt, D., George, J. M., Biegler, M. T., Iorns, D., Digby, A., Eason, D., Robertson, B., Edwards, T., Wilkinson, M., Turner, G., Meyer, A., Kautt, A. F., Franchini, P., Detrich, H. W., 3rd, Svardal, H., Wagner, M., Naylor, G. J. P., Pippel, M., Malinsky, M., Mooney, M., Simbirsky, M., Hannigan, B. T., Pesout, T., Houck, M., Misuraca, A., Kingan, S. B., Hall, R., Kronenberg, Z., Sović, I., Dunn, C., Ning, Z., Hastie, A., Lee, J., Selvaraj, S., Green, R. E., Putnam, N. H., Gut, I., Ghurye, J., Garrison, E., Sims, Y., Collins, J., Pelan, S., Torrance, J., Tracey, A., Wood, J., Dagnew, R. E., Guan, D., London, S. E., Clayton, D. F., Mello, C. V., Friedrich, S. R., Lovell, P. V., Osipova, E., Al-Ajli, F. O., Secomandi, S., Kim, H., Theofanopoulou, C., Hiller, M., Zhou, Y., Harris, R. S., Makova, K. D., Medvedev, P., Hoffman, J., Masterson, P., Clark, K., Martin, F., Howe, K., Flicek, P., Walenz, B. P., Kwak, W., Clawson, H., Diekhans, M., Nassar, L., Paten, B., Kraus, R. H. S., Crawford, A. J., Gilbert, M. T. P., Zhang, G., Venkatesh, B., Murphy, R. W., Koepfli, K.-P., Shapiro, B., Johnson, W. E., Di Palma, F., Marques-Bonet, T., Teeling, E. C., Warnow, T., Graves, J. M., Ryder, O. A., Haussler, D., O'Brien, S. J., Korlach, J., Lewin, H. A., Howe, K., Myers, E. W., Durbin, R., Phillippy, A. M. & Jarvis, E. D. Towards complete and error-free genome assemblies of all vertebrate species. *Nature* **592,** 737–746 (2021).

35. Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9,** 559 (2008).

36. Gedman, G., Haase, B., Durieux, G., Biegler, M. T., Fedrigo, O. & Jarvis, E. D. 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

37. 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).

38. Teramitsu, I. & White, S. A. FoxP2 regulation during undirected singing in adult songbirds. *J. Neurosci.* **26,** 7390–7394 (2006).

39. Kim, J., Lee, C., Ko, B. J., Yoo, D., Won, S., Phillippy, A., Fedrigo, O., Zhang, G., Howe, K., Wood, J., Durbin, R., Formenti, G., Brown, S., Cantin, L., Mello, C. V., Cho, S., Rhie, A., Kim, H. & Jarvis, E. D. False gene and chromosome losses affected by assembly and sequence errors. *bioRxiv* 2021.04.09.438906 (2021). doi:10.1101/2021.04.09.438906

40. Itoh, Y. & Arnold, A. P. Chromosomal polymorphism and comparative painting analysis in the zebra finch. *Chromosome Res.* **13,** 47–56 (2005).

41. Warren, W. C., Clayton, D. F., Ellegren, H., Arnold, A. P., Hillier, L. W., Künstner, A., Searle, S., White, S., Vilella, A. J., Fairley, S., Heger, A., Kong, L., Ponting, C. P., Jarvis, E. D., Mello, C. V., Minx, P., Lovell, P., Velho, T. A. F., Ferris, M., Balakrishnan, C. N., Sinha, S., Blatti, C., London, S. E., Li, Y., Lin, Y.-C., George, J., Sweedler, J., Southey, B., Gunaratne, P., Watson, M., Nam, K., Backström, N., Smeds, L., Nabholz, B., Itoh, Y., Whitney, O., Pfenning, A. R., Howard, J., Völker, M., Skinner, B. M., Griffin, D. K., Ye, L., McLaren, W. M., Flicek, P., Quesada, V., Velasco, G., Lopez-Otin, C., Puente, X. S., Olender, T., Lancet, D., Smit, A. F. A., Hubley, R., Konkel, M. K., Walker, J. A., Batzer, M. A., Gu, W., Pollock, D. D., Chen, L., Cheng, Z., Eichler, E. E., Stapley, J., Slate, J., Ekblom, R., Birkhead, T., Burke, T., Burt, D., Scharff, C., Adam, I., Richard, H., Sultan, M., Soldatov, A., Lehrach, H., Edwards, S. V., Yang, S.-P., Li, X., Graves, T., Fulton, L., Nelson, J., Chinwalla, A., Hou, S., Mardis, E. R. & Wilson, R. K. The genome of a songbird. *Nature* **464,** 757–762 (2010).

42. Itoh, Y., Melamed, E., Yang, X., Kampf, K., Wang, S., Yehya, N., Van Nas, A., Replogle, K., Band, M. R., Clayton, D. F., Schadt, E. E., Lusis, A. J. & Arnold, A. P. Dosage compensation is less effective in birds than in mammals. *J. Biol.* **6,** 2 (2007).

43. Appenzeller, S., Schirmacher, A., Halfter, H., Bäumer, S., Pendziwiat, M., Timmerman, V., De Jonghe, P., Fekete, K., Stögbauer, F., Lüdemann, P., Hund, M., Quabius, E. S., Bernd Ringelstein, E. & Kuhlenbäumer, G. Autosomal-Dominant Striatal Degeneration Is Caused by a Mutation in the Phosphodiesterase 8B Gene. *The American Journal of Human Genetics* **86,** 83–87 Preprint at https://doi.org/10.1016/j.ajhg.2009.12.003 (2010)

44. 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).

45. Chen, Y.-C., Kuo, H.-Y., Bornschein, U., Takahashi, H., Chen, S.-Y., Lu, K.-M., Yang, H.-Y., Chen, G.-M., Lin, J.-R., Lee, Y.-H., Chou, Y.-C., Cheng, S.-J., Chien, C.-T., Enard, W., Hevers, W., Pääbo, S., Graybiel, A. M. & Liu, F.-C. Foxp2 controls synaptic wiring of corticostriatal circuits and vocal communication by opposing Mef2c. *Nat. Neurosci.* **19,** 1513–1522 (2016).
46. 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).
47. 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).
48. Harris, S. E., Cox, S. R., Bell, S., Marioni, R. E., Prins, B. P., Pattie, A., Corley, J., Muñoz Maniega, S., Valdés Hernández, M., Morris, Z., John, S., Bronson, P. G., Tucker-Drob, E. M., Starr, J. M., Bastin, M. E., Wardlaw, J. M., Butterworth, A. S. & Deary, I. J. Neurology-related protein biomarkers are associated with cognitive ability and brain volume in older age. *Nat. Commun.* **11,** 800 (2020).
49. Ruth, K. S., Day, F. R., Tyrrell, J., Thompson, D. J., Wood, A. R., Mahajan, A., Beaumont, R. N., Wittemans, L., Martin, S., Busch, A. S., Erzurumluoglu, A. M., Hollis, B., O'Mara, T. A., Endometrial Cancer Association Consortium, McCarthy, M. I., Langenberg, C., Easton, D. F., Wareham, N. J., Burgess, S., Murray, A., Ong, K. K., Frayling, T. M. & Perry, J. R. B. Using human genetics to understand the disease impacts of testosterone in men and women. *Nat. Med.* **26,** 252–258 (2020).
50. Kornfeld, M., Synder, R. D., MacGee, J. & Appenzeller, O. The oculo-cerebral-renal syndrome of Lowe. *Arch. Neurol.* **32,** 103–107 (1975).
51. Ates, K. M., Wang, T., Moreland, T., Veeranan-Karmegam, R., Ma, M., Jeter, C., Anand, P., Wenzel, W., Kim, H.-G., Wolfe, L. A., Stephen, J., Adams, D. R., Markello, T., Tifft, C. J., Settlage, R., Gahl, W. A., Gonsalvez, G. B., Malicdan, M. C., Flanagan-Steet, H. & Pan, Y. A. Deficiency in the endocytic adaptor proteins PHETA1/2 impairs renal and craniofacial development. *Dis. Model. Mech.* **13,** (2020).

52. 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).

53. Gao, J., Kang, X.-Y., Sun, S., Li, L., Zhang, B.-L., Li, Y.-Q. & Gao, D.-S. Transcription factor Six2 mediates the protection of GDNF on 6-OHDA lesioned dopaminergic neurons by regulating Smurf1 expression. *Cell Death Dis.* **7,** e2217 (2016).

54. 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).

55. 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).

56. 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).

57. Friedrich, S. R., Nevue, A. A., Andrade, A. L. P., Velho, T. A. F. & Mello, C. V. Emergence of sex-specific transcriptomes in a sexually dimorphic brain nucleus. *Cell Rep.* **40,** 111152 (2022).

58. 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).

59. 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).

60. De Vries, G. J., Rissman, E. F., Simerly, R. B., Yang, L.-Y., Scordalakes, E. M., Auger, C. J., Swain, A., Lovell-Badge, R., Burgoyne, P. S. & Arnold, A. P. A model system for study of sex chromosome effects on sexually dimorphic neural and behavioral traits. *J. Neurosci.* **22,** 9005–9014 (2002).

61. 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).

62. 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).

63. 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).

64. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30,** 2114–2120 (2014).

65. Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M. & Gingeras, T. R. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29,** 15–21 (2013).

66. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30,** 923–930 Preprint at https://doi.org/10.1093/bioinformatics/btt656 (2014)

67. 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).

68. Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T., Miller, E., Bache, S., Müller, K., Ooms, J., Robinson, D., Seidel, D., Spinu, V., Takahashi, K., Vaughan, D., Wilke, C., Woo, K. & Yutani, H. Welcome to the tidyverse. *J. Open Source Softw.* **4,** 1686 (2019).

69. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*. (Springer Science & Business Media, 2009).

70. 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).

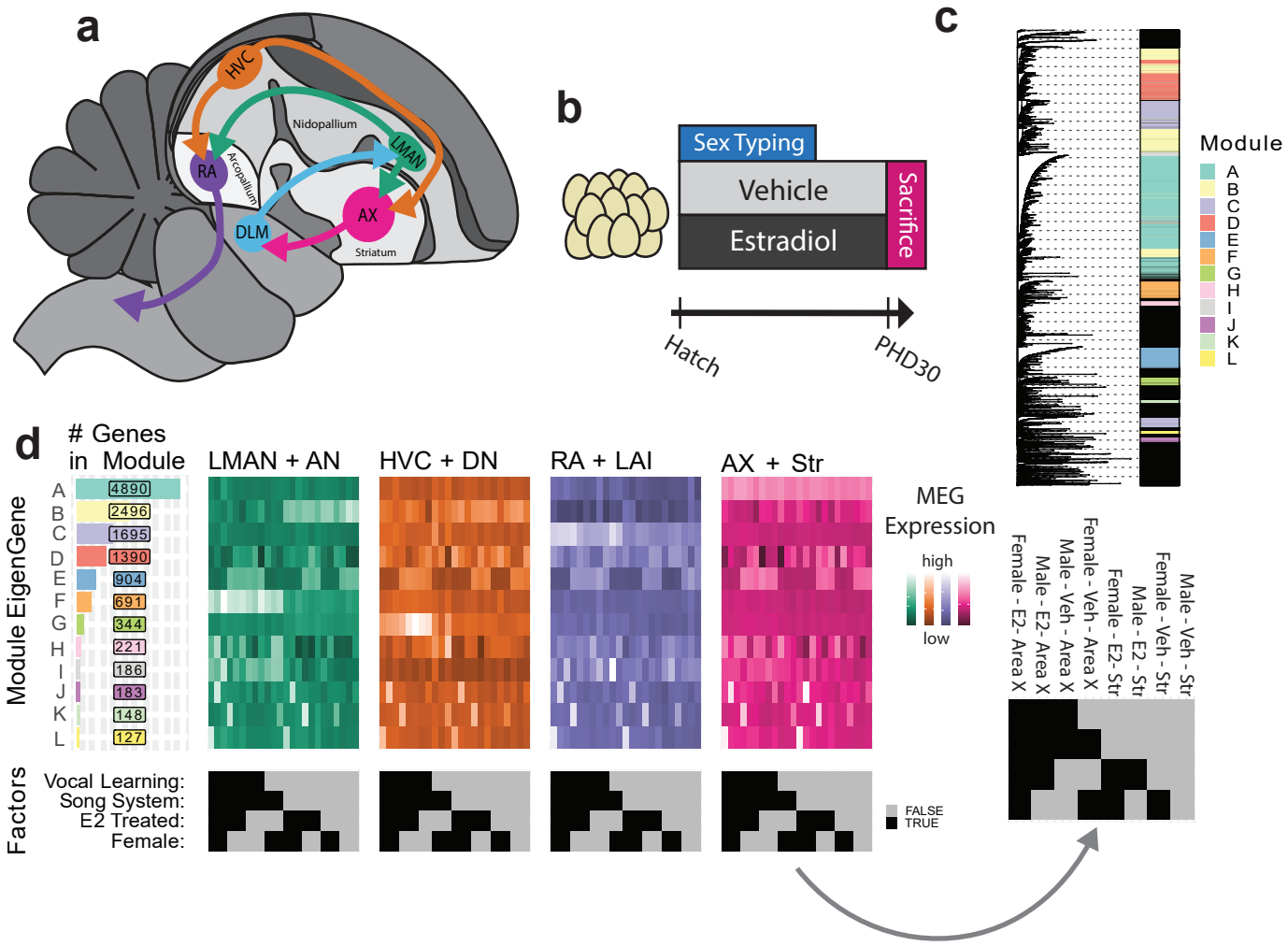

Figure 1 - Song system anatomy and experimental design. a, Diagram of song system connectivity within the adult male zebra finch brain with major telencephalic domains indicated. Area X connects back to LMAN through the non-vocal specific thalamic nucleus DLM. **b,** Experimental design. Animals were treated with E2 or a vehicle from hatch until sacrifice on PHD30. **c,** WGCNA assignment of genes to modules. Left: Hierarchy computed over the transcriptome wide topological overlap matrix of gene to gene correlations in transcript abundance across samples. Right: Module assignment raster, rows are genes colored according to the assigned module, unassigned genes in black. **d,** MEG expression heatmaps arranged by module size (left) aligned to traits of interest (bottom). Each row is an MEG and each sample is a column. Samples are grouped according to neural circuit node in different colored subpanels. Color intensity encodes MEG expression as calculated by WGCNA. An example raster with sample category labels is provided at right.

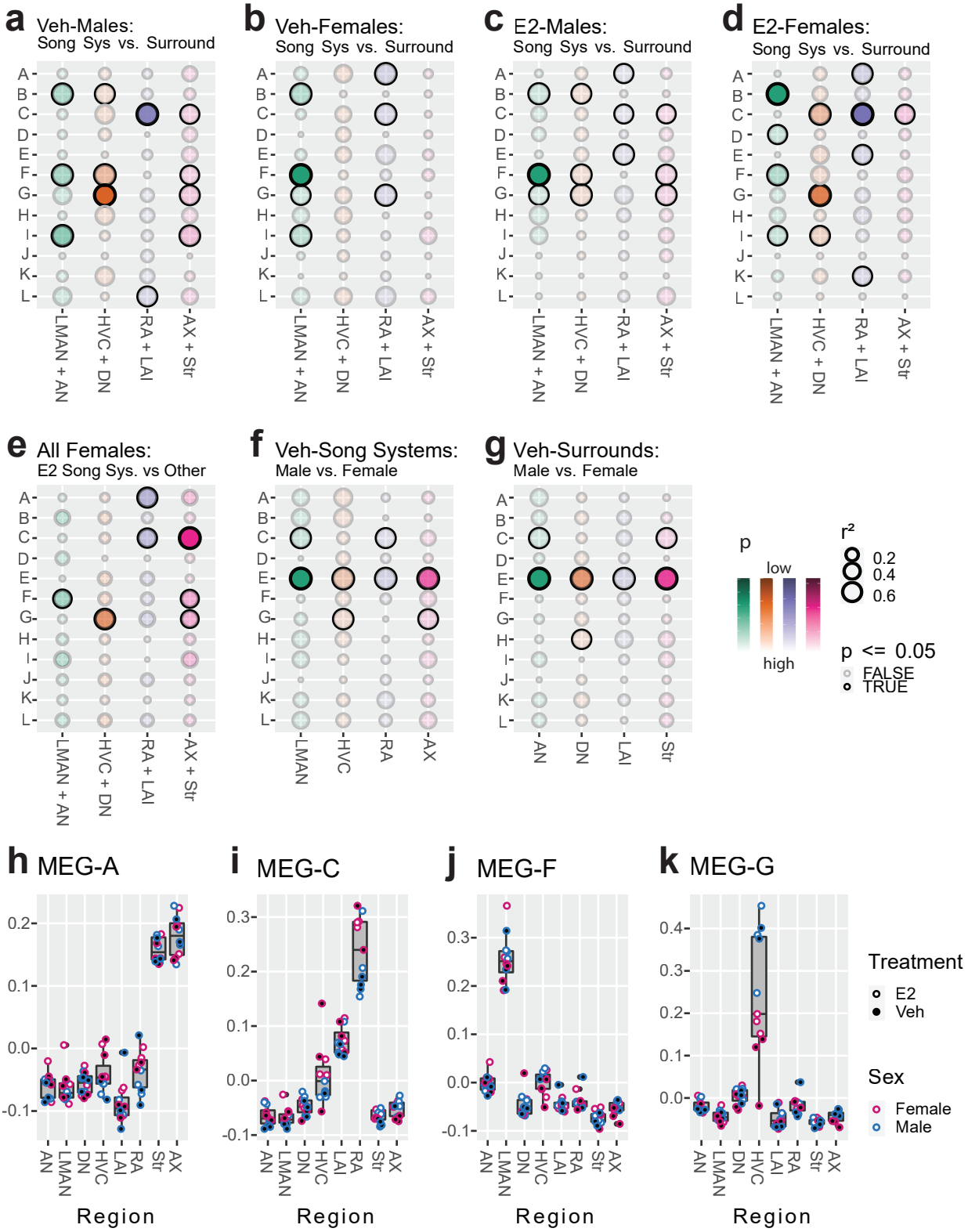
a Veh-Males:
Song Sys vs. Surround
b Veh-Females:
Song Sys vs. Surround
c E2-Males:
Song Sys vs. Surround
d E2-Females:
Song Sys vs. Surround
A
B
C
D
E
F
G
H
I
J
K
L
LMAN + AN
HVC + DN
RA + LAI
AX + Str
e All Females:
E2 Song Sys. vs Other
f Veh-Song Systems:
Male vs. Female
g Veh-Surrounds:
Male vs. Female
LMAN
HVC
RA
AX
AN
DN
LAI
Str
p
low
high
r²
0.2
0.4
0.6
p <= 0.05
FALSE
TRUE
h MEG-A
i MEG-C
j MEG-F
k MEG-G
Region
Treatment
E2
Veh
Sex
Female
Male

Figure 2 - Association of modules to experimental variables. a-g, Bubble plots showing statistical association between MEG expression and variables of interest in various sample subsets. Strength of association (r^2) is encoded by bubble size, significance (p) is encoded in the color scale with significant associations darkly bordered. Pearson correlation and Students t test, alpha=0.05. Plots show the associations between gene modules (rows) to; **a-b,** vehical treated song system specializations, comparing MEG expression in song system samples from either sex to their appropriate surrounding controls; **c-d,** E2-treated song system specialization, same comparison as **a-b** but within E2-treated samples; **e,** female vocal learning capacity after E2, comparing E2 treated female song system components to all other female samples from that circuit node; **f,** sexual dimorphism within the song system, comparing vehicle treated male and female song system components; **g,** sexual dimorphism within the surrounding control regions, comparing the vehicle treated male and female surrounding control samples. Each neural circuit node is considered separately (columns). **h-k,** Expression of modules with strong region specific expression. Module A is highly expressed in Str and Area X samples, with addtional differences between RA and LAI (**h**); Module C is highly expressed in the arcopallium, especially RA, with some increaase in HVC (**i**); Module F is expressed highly only in LMAN regardless of sex or treatement (**j**); Module G is only highly expressed in HVC, where it differs both by sex and treatment (**k**).

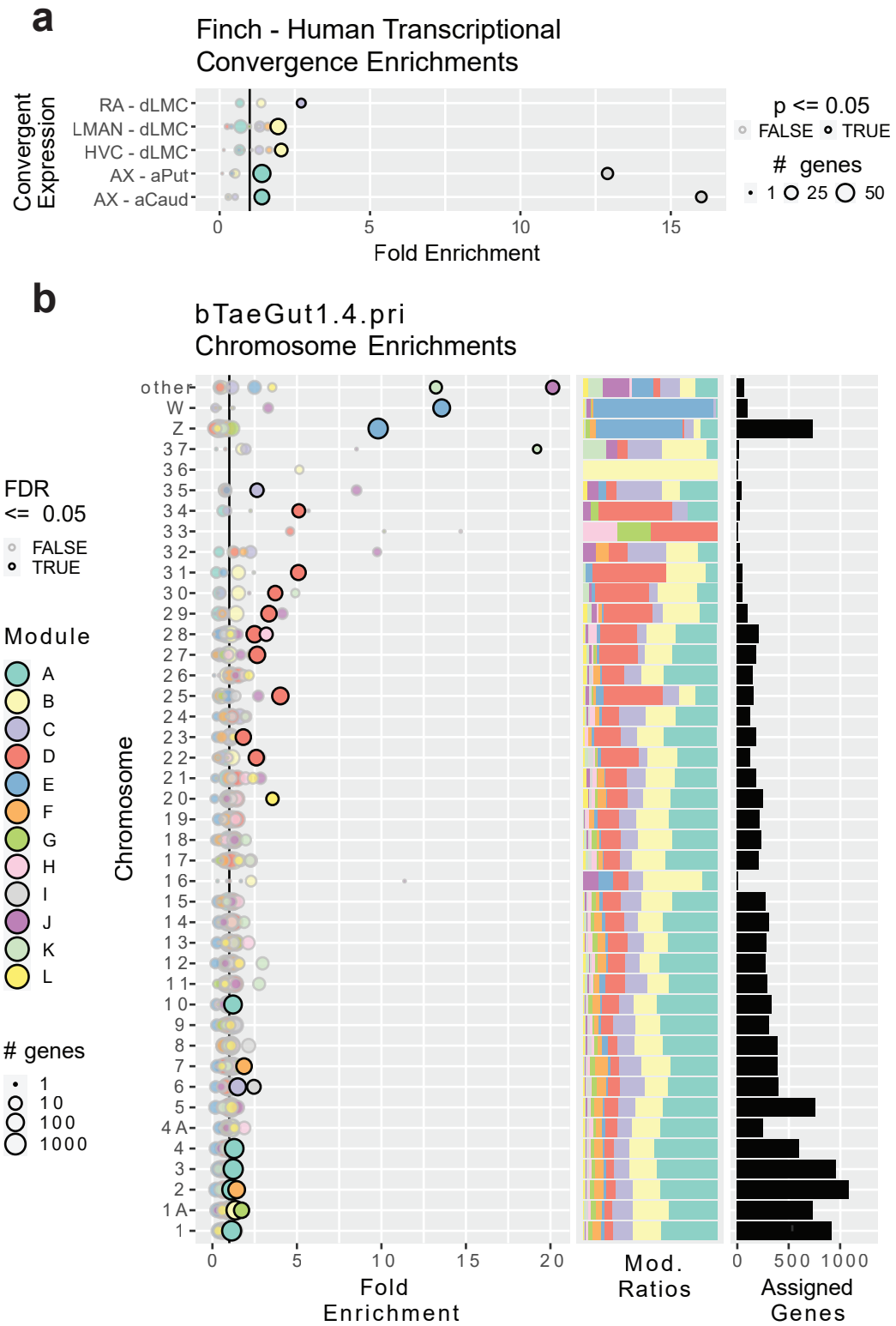

Figure 3 - Gene module enrichments for human convergent signature and for chromosomes. a, Enrichment of genes previously found to be convergently differentially expressed in the human laryngeal motor cortex and the pallial song nuclei or convergently expressed between the human vocal striatum and Area X. Bubble size linearly encodes the number of genes in each convergence module pairing. Significance was assed using a one-tailed GAGE test, similar to GO ontologies, alpha=0.05. Significant enrichments are darkly bordered and opaque. Values to the right of the vertical black line indicate above random chance. **b,** Enrichment of genes from specific chromosomes. Left**,** fold enrichment of modules onto zebra finch chromosomes in the newest genome assembly available; center, the portion of module assigned trascripts from each chromosome per module; right, the number of module assinged genes per chromosome. Each row is a chromosome with each bubble representing the enrichment of transcripts from that chromosome in one of the gene module defined by WGNCA. Values to the right of the vertical black line indicate above random chance. The size of the bubbles indicates the log10 transformed number of genes in each chromosome module pairing. Significance was assessed using an FDR corrected bootstrapped test of observed enrichment for each module chromosome pairing based on 50,000 randomizations of genes into modules. Significant enrichments are darkly bordered and opaque. **a-b** use the same color scale for modules.

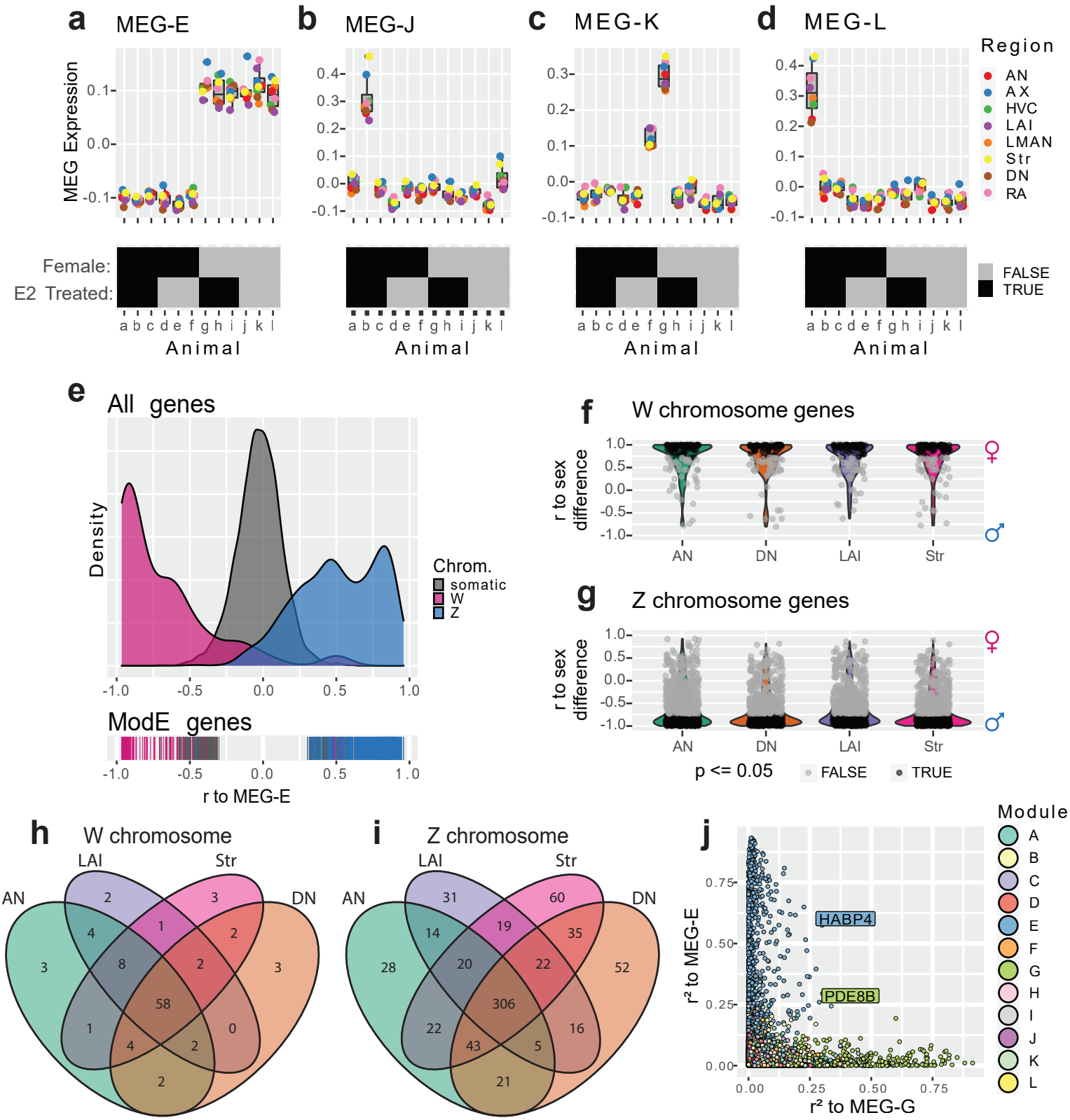
a
MEG-E
b
MEG-J
c
MEG-K
d
MEG-L
MEG Expression
Region
AN
AX
HVC
LAI
LMAN
Str
DN
RA
Female:
E2 Treated:
FALSE
TRUE
Animal
e
All genes
Density
Chrom.
somatic
W
Z
ModE genes
r to MEG-E
f
W chromosome genes
r to sex difference
g
Z chromosome genes
p <= 0.05
h
W chromosome
i
Z chromosome
j
Module
HABP4
PDE8B
r² to MEG-E
r² to MEG-G

Figure 4 - Brainwide signatures of sex and micro chromosome expression. a-d, Expression of selected module eigengenes by animal (top) aligned to their respective experimental variables (bottom), color indicates region. The sex chromosome enriched module E was highly expressed in all male samples and depleted in all female samples regardless of brain region or pharmacological treatment (a). Module J, K, and L eigengenes were each highly expressed in samples from one (J and L) or two (K) animals across all brain regions sampled. **e,** Distribution of continuous membership in module E across all module assigned genes (top) and module E assigned genes (bottom) based on corelation of expression to the module eigengene (Pearson r to MEG-E) with sex chromosomes separated. **f-g**, Distribution of sex chromosome gene expression correlations to the sex difference in vehicle treated finches. Postive correlations indicate female biased expression while anticorrelations indicate male biased expression. Significance was assessed in each region using an upper-tailed student's correlation test for W chromosome transcripts (**f**) and lower-tailed for Z chromosome transcripts (**g**), with significant correlations in black, alpha=0.05. **h-i,** Venn diagrams intersecting the significantly sex difference correlated genes across non-vocal surround regions for the W and Z chromosomes respectively. **j,** Comparison of continuous membership in module E (r2 to MEG-E, y-axis) and module G (r2 to MEG-G, x-axis) across all

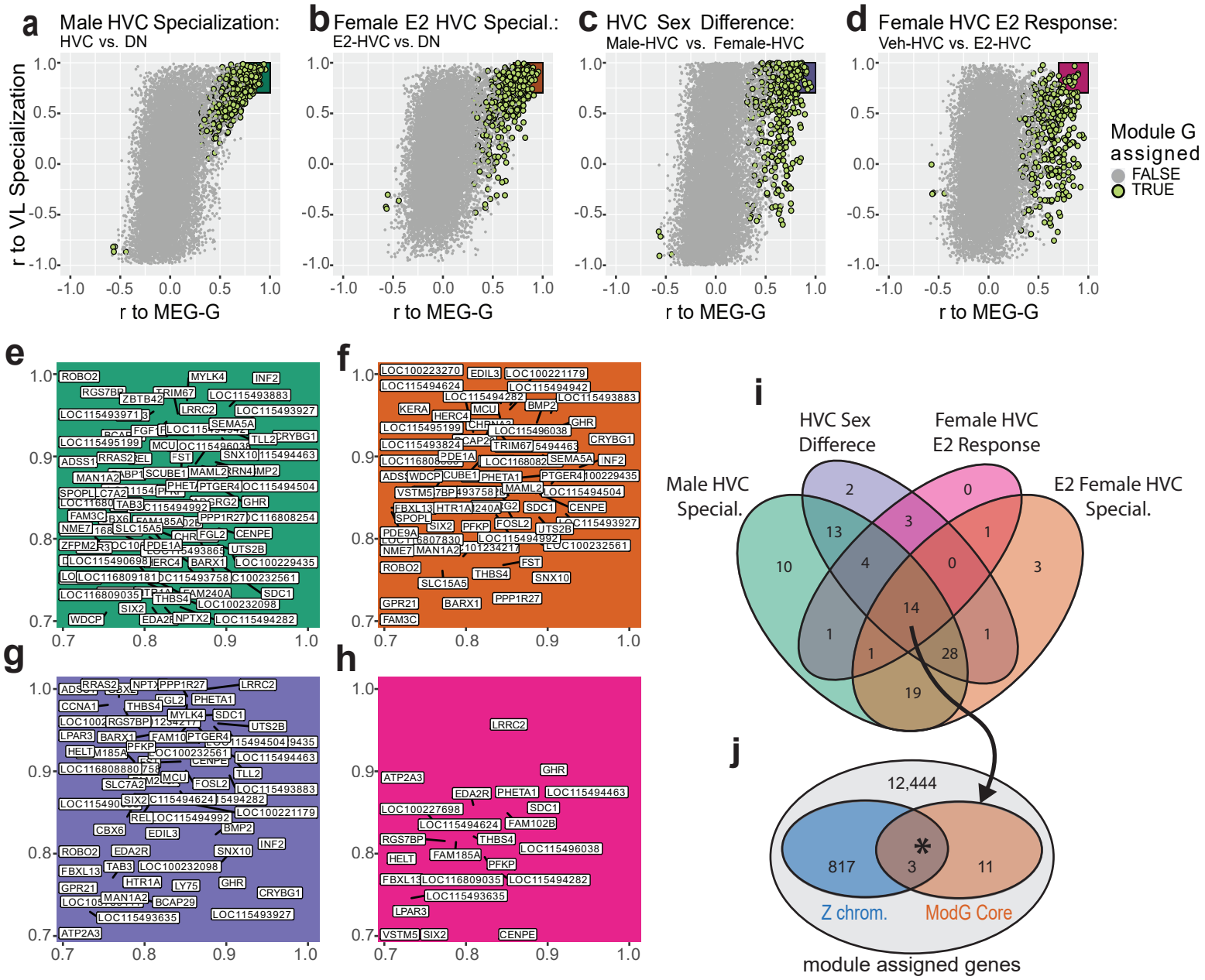

Figure 5 - Identification of core genes in module G and their association to the Z chromosome. a-d, Single gene continuous membership in module G (x-axis; Pearson r to MEG from module G) for all assigned genes vs correlation to vocal learning in masculine or masculinized HVC relative to samples from non-vocal learning females in each of the four comparisons; **a,** male song system membership, comparing individual gene expression in male HVC samples of either treatment to expression in the surrounding DN; **b,** female vocal learning capacity after E2, comparing E2 treated HVC to all other female DN or HVC samples; **c,** sexual dimorphic gene expression within the song system, comparing vehicle treated male and female song system components; **d,** estradiol responsive gene expression in female HVC, comparing E2 treat and vehicle treat female HVC samples. Each point is a gene colored by module assignment, the shaded area indicates gene of interest criteria for each comparison. **e-h,** Blowup of shaded regions in **a-d** respectively showing genes of interest from each comparison. **i,** Identification of core genes by intersecting the four gene sets of interest. **j,** Enrichment of Z chromosome transcripts within the core genes. * indicates p = 0.0087 by an upper-tailed hypergeometric test.

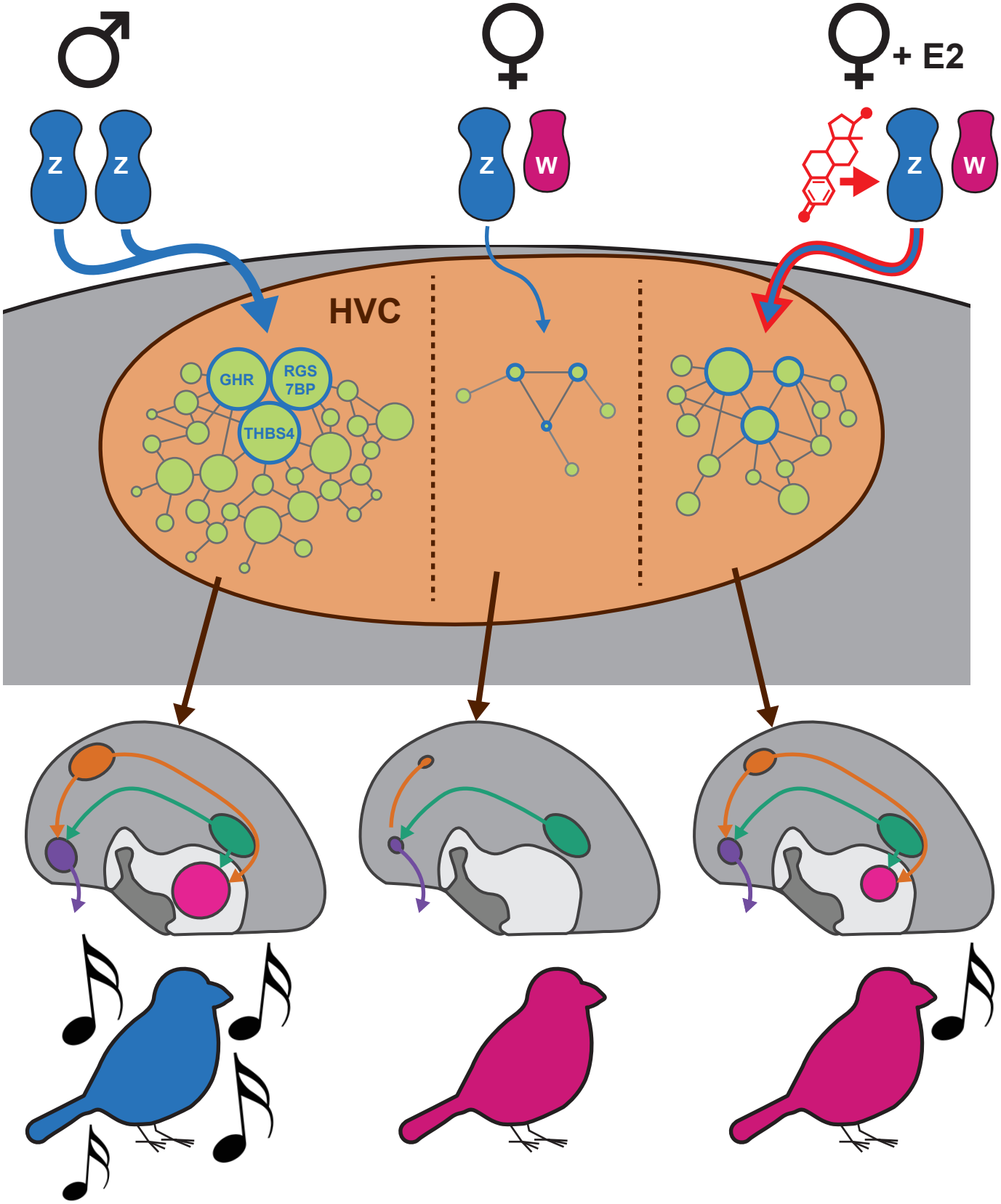

Figure 6 - Proposed model of sexually dimorphic zebra finch vocal learning. We propose that estradiol treatment in female zebra finches masculinizes song behavior by overcoming insufficient Z sex chromosome dosage in HVC to increase the expressed of transcripts normally depleted in females. The Z chromosome genes upregulated by E2 are components in a larger proliferative genetic program which prevents HVC atrophy in males and allows for its expansion late in development. The upregulation of these genes allows for the increased specilization of the gene networks they participate in, promoting HVC development sufficiently to enable 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 uncorrected p value calculated by GAGE.

Module E: Significantly Enriched GO

GOterm	GOid	p
nucleolus	GO:0005730	0.024
transcription, DNA-templated	GO:0006351	0.031
DNA repair	GO:0006281	0.037
damaged DNA binding	GO:0003684	0.044
U2-type precatalytic spliceosome	GO:0071005	0.047
cellular response to DNA damage stimulus	GO:0006974	0.048

Table 2 - Module E functional enrichment analysis. Significantly enriched GO terms within the 1:1 human orthologs of module G. Lists full GO term, GOid, and uncorrected p value calculated by GAGE.

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 3 - 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.

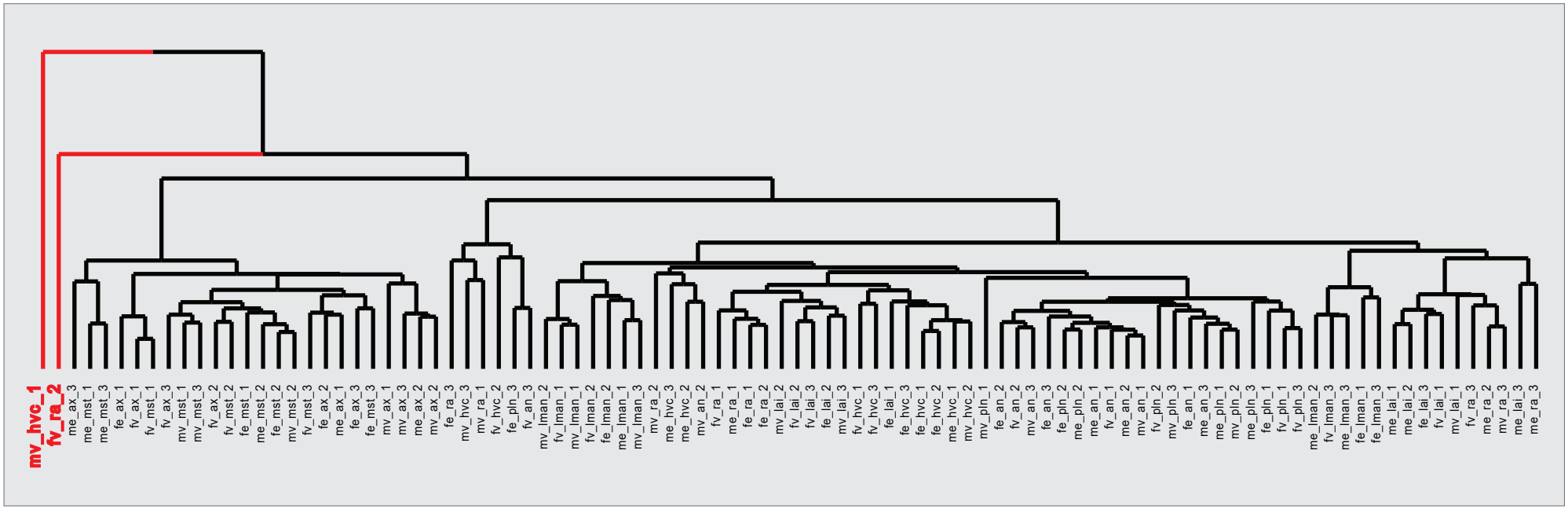

Figure S1 - Outlier sample detection by hierarchical clustering. Two samples (a vehicle-treated male HVC sample and an E2-treated female RA sample, in red) form single sample branches in the hierarchical clustering tree, indicative of technical outliers unlikely to fit the correlational structure of the larger dataset. Samples were removed prior to gene network construction and module detection. For clustering of included samples, see **Fig S5**.

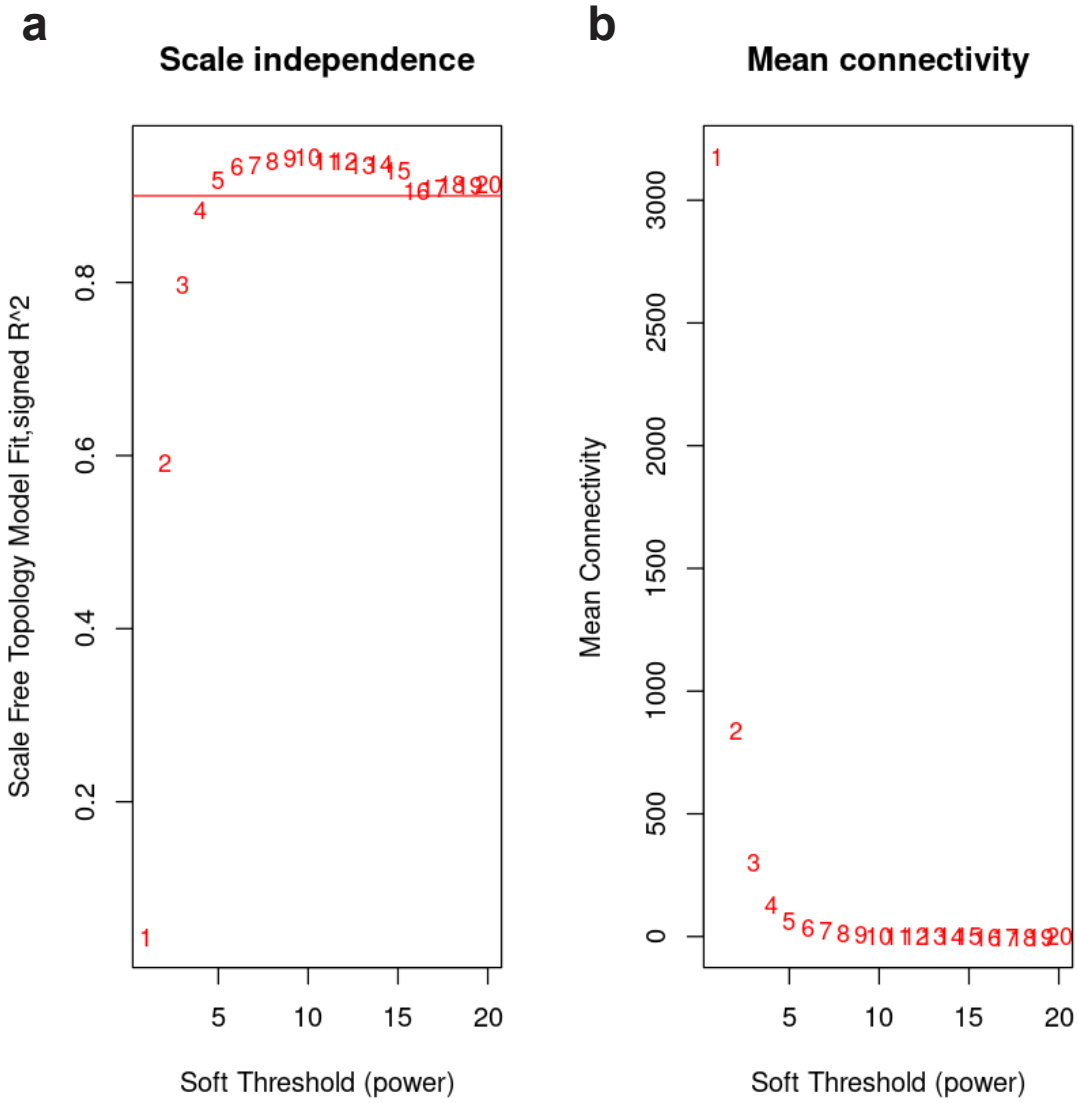

Figure S2 - Selection of soft thresholding power for WGCNA model. Soft-thresholding power (beta, x-axis) is the exponent to which each element in the gene-to-gene correlation matrix is raised during adjacency matrix calculation. **a,** Scale-free fit index (y-axis) as a function of the soft-thresholding power (x-axis). Horizontal line indicates a fit of 90%. **b,** Mean connectivity (degree) in the network model (y-axis) as a function of the soft-thresholding power. We selected a power of 6 as it is on the knee of both plots and above the 90% scale free fit criteria.

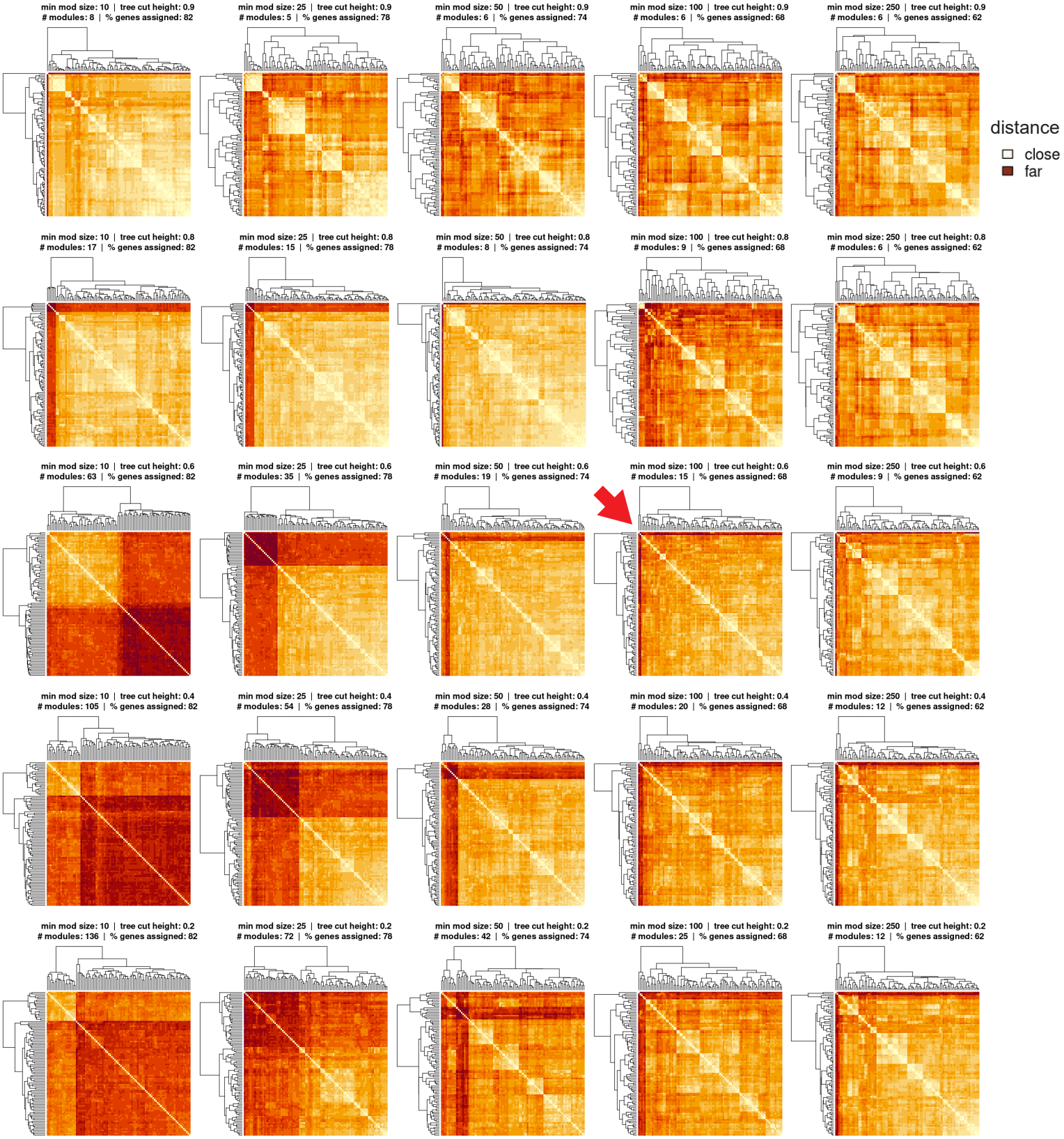

Figure S3 - Selection of minimum module size and tree cut height parameter values for WGCNA model. Each plot shows the sample distance matrix that results from the parameters in the plot title. Titles also show the number of modules found in each resulting model and the percentage of genes in the finch genome assigned. Minimum module size increases across columns (left to right: 10, 25, 50, 100, 250) and tree cut height decreases down rows (top to bottom: 0.9, 0.8, 0.6, 0.4, 0.2). Black arrow indicates selected model. Model was selected to explain as much transcriptomic variance as possible while minimizing the number of technically overfit samples.

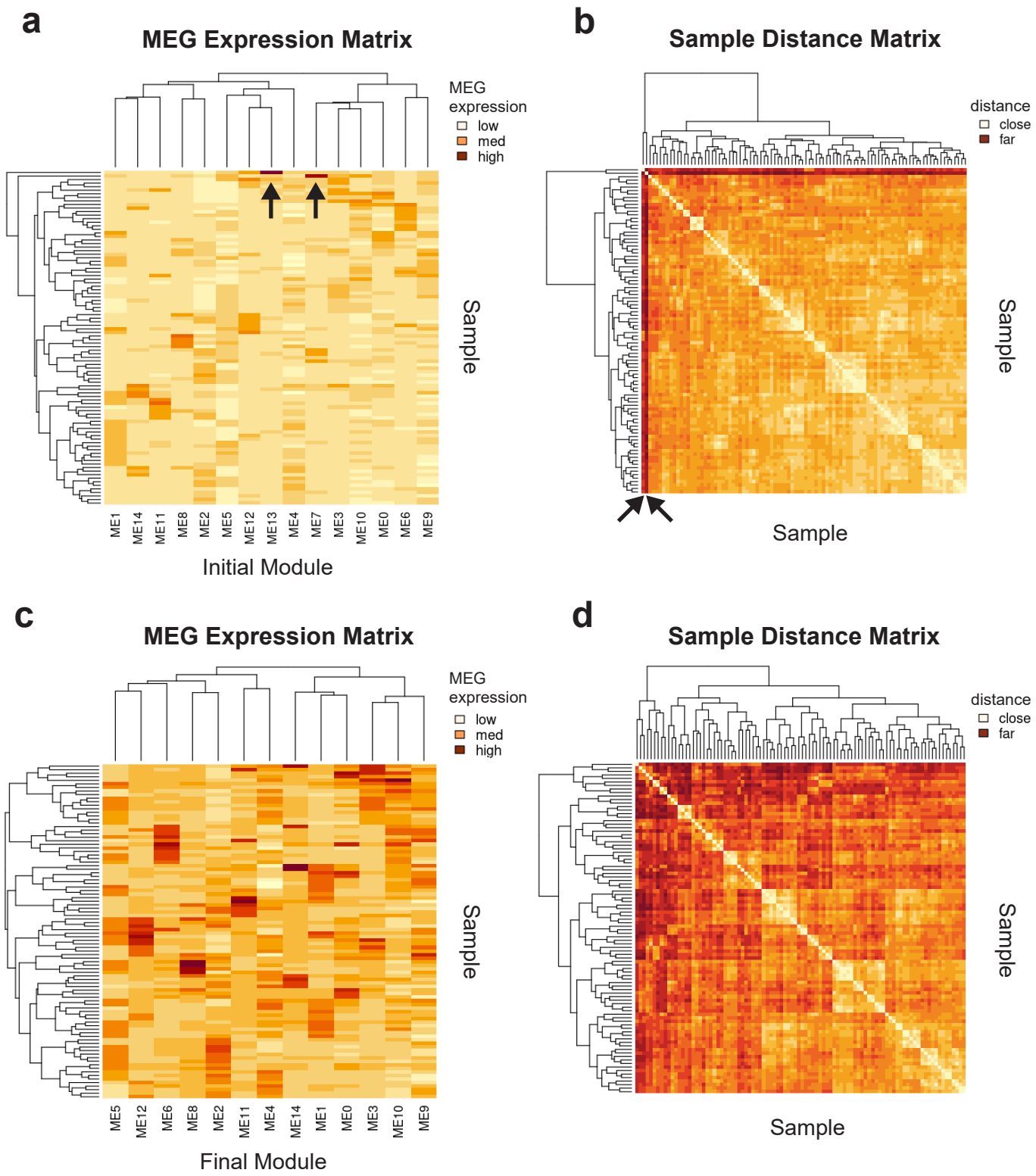

Figure S4 - Initial module overfitting to single samples. a, Module eigengene (MEG) 7 and 13 are both highly expressed only in sinlge samples, indicated by black arrows. **b,** This overfitting causes these samples to be deep outliers in the sample-sample distance matrix, distant from all samples but tthemselves, indicated by black arrows. **c,** Removing these module eigengenes from the set prevents these samples from behaving as outliers in the distance matrix (**d**). These overfit modules were removed prior to module lettering and statistical analysis.

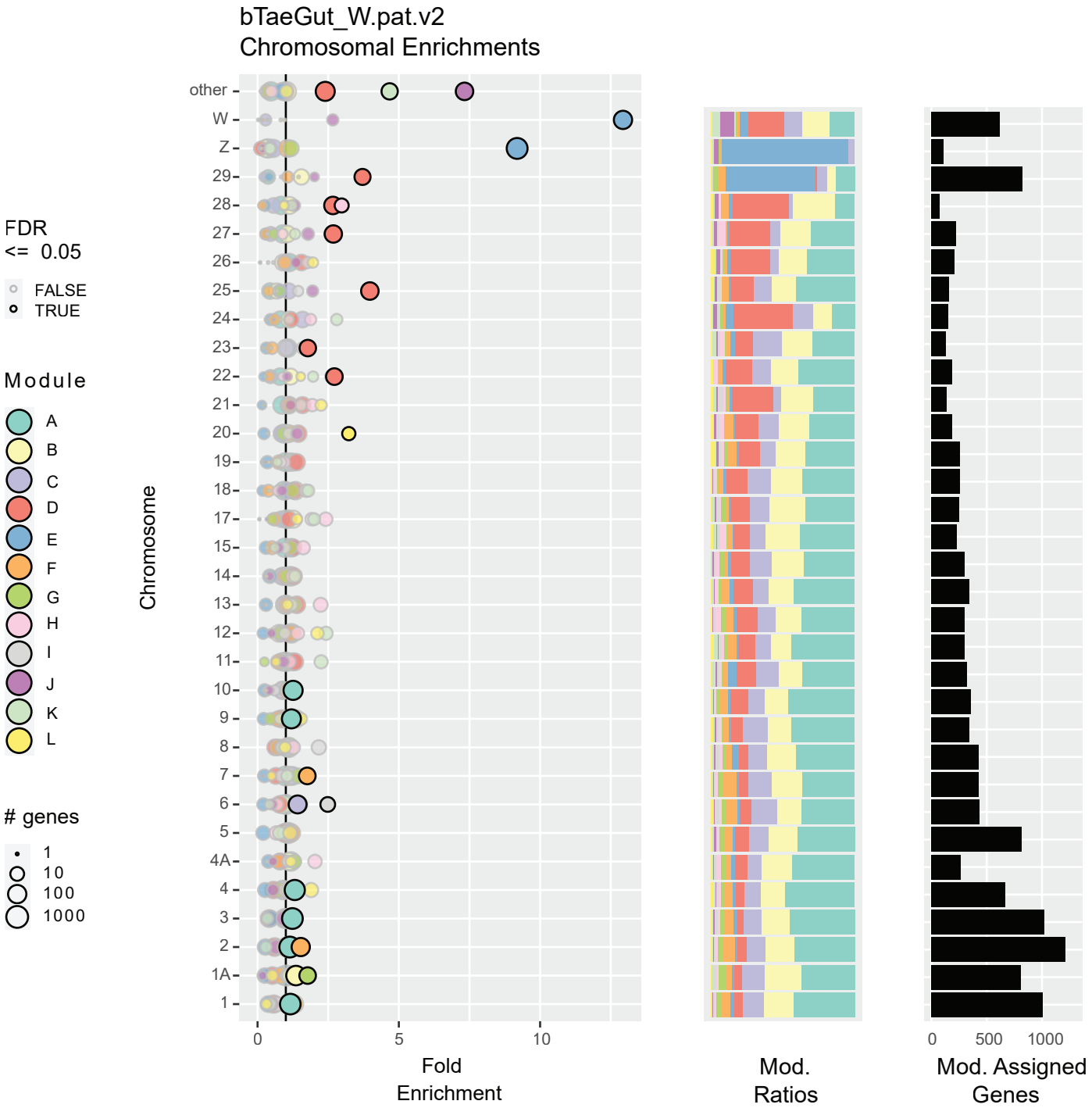

Figure S5 - Gene module enriched for chromosomes and for human convergent signature. Left, fold enrichment of modules onto zebra finch chromosomes in the newest genome assembly available; center, the portion of module assigned trascripts from each chromosome per module; right, the number of module assinged genes per chromosome. Each row is a chromosome with each bubble representing the enrichment of transcripts from that chromosome in one of the gene module defined by WGNCA. Values to the right of the vertical black line indicate above random chance. The size of the bubbles indicate the number of genes in each chromosome module pairing. Significance was assessed using an FDR corrected bootstrapped test of observed enrichment for each module chromosome pairing based on 50,000 randomizations of genes into modules. Significant enrichments are darkly bordered and opaque.

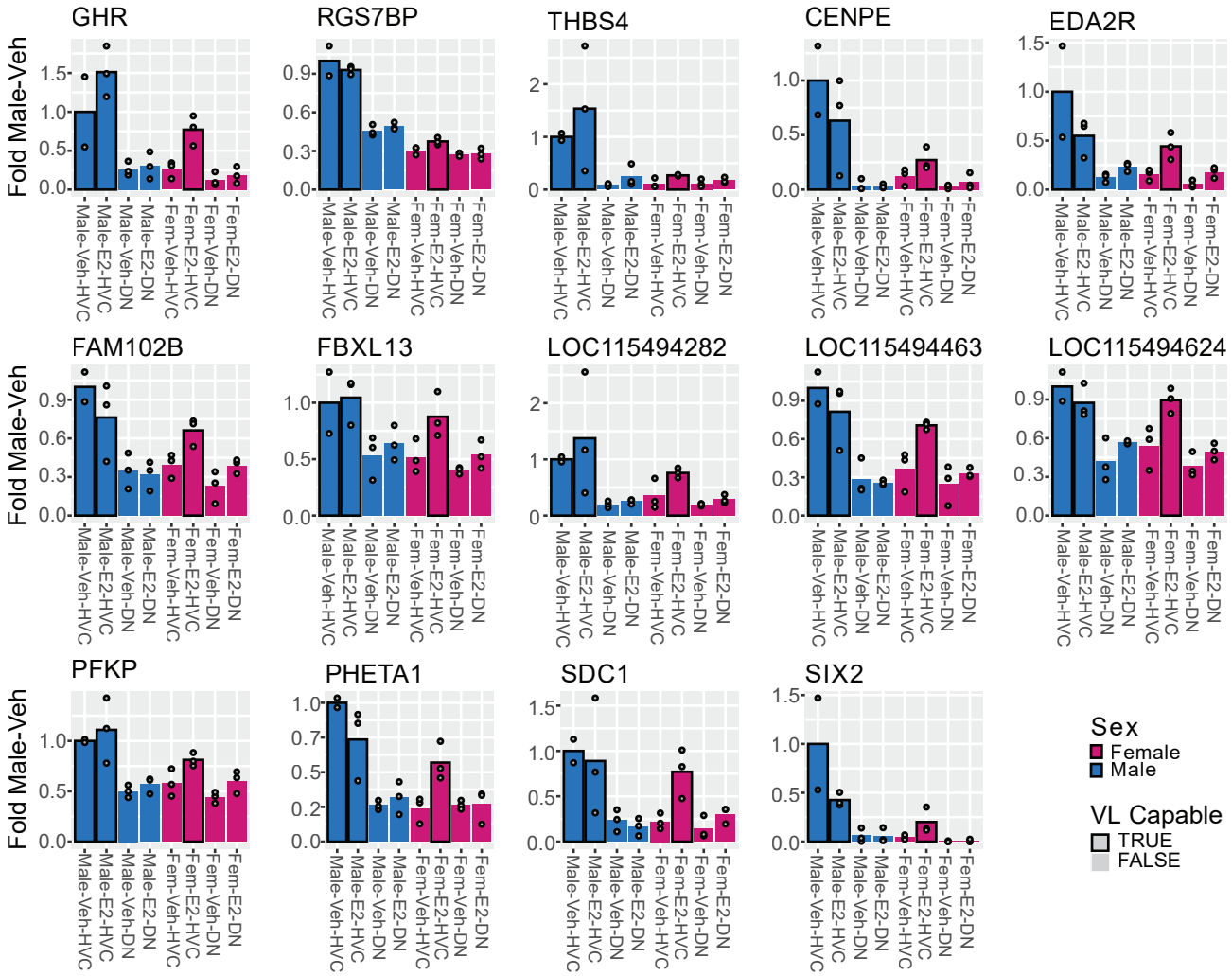

Figure S6 - Expression of module G core genes in HVC and surrounding dorsal nidopallium. Each of the 14 core genes show reduced expression in female HVC relative to the male with an increase in expression in response to E2 treatment. Bar represents mean with individual data points shown. This transcriptional response to E2 is not seen in 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

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

Z chromosome								W chromosome	
ABHD17B	CPLANE1	FOCAD	LOC100224449	MED18	PLIN2	SERF1A	TMEM171	CHD1W	LOC116806914
ACAA2	CREB3	FSD1L	LOC100224709	MIER3	PLPP1	SETBP1	TMEM267	LOC100189940	LOC116806917
ACO1	CSNK1G3	FUT10	LOC100225594	MIR2954	POLR1E	SETD9	TMEM38B	LOC100190109	LOC116806918
AGGF1	CWC27	FXN	LOC100226145	MOCS2	PPIC	SGTB	TNPO1	LOC100190380	LOC116806920
AGTPBP1	CZH18orf25	GALT	LOC100226207	MPDZ	PPIP5K2	SHC3	TOPORS	LOC115492298	LOC116806923
AK3	CZH18orf32	GIN1	LOC100229438	MRPL17	PPWD1	SHLD3	TPGS2	LOC116806641	LOC116806940
AMACR	CZH5orf51	GNE	LOC100230245	MRPS27	PSIP1	SLC30A5	TRAPPC13	LOC116806651	LOC116806947
ANKRA2	CZH5orf63	GNG10	LOC100230481	MRPS30	PTAR1	SLC38A9	TRIM23	LOC116806853	LOC116806950
ANKRD55	CZH9orf40	GOLM1	LOC105760834	MRPS36	PTBP3	SLC44A1	TSTD2	LOC116806854	LOC116806951
AOPEP	CZH9orf64	GOLPH3	LOC105760853	MTAP	PTCD2	SLF1	TTC33	LOC116806857	LOC116806966
AP3B1	CZH9orf85	GRIN3A	LOC115490934	MTREX	PTGR1	SMAD2	TTC37	LOC116806862	LOC116806973
APBA1	DCP2	HABP4	LOC115490942	MTX3	PUM3	SMAD7	TUSC1	LOC116806863	LOC116806978
APTX	DCTN3	HACD4	LOC115490943	MYORG	RAB3C	SMARCA2	TUT7	LOC116806864	LOC116806979
ARHGEF39	DDX58	HAUS1	LOC115490975	NAA35	RAD1	SMC5	TXN	LOC116806866	LOC116806985
ARRDC3	DHX29	HAUS6	LOC115490991	NDUFAF2	RAD17	SMIM15	UBAP1	LOC116806867	LOC116806986
ATG12	DNAJA1	HEXB	LOC115491002	NDUFS4	RASA1	SNAPC3	UBAP2	LOC116806869	LOC116806988
ATP5F1A	DNAJC21	HINT1	LOC115491073	NFIB	RCL1	SNCAIP	UBE2R2	LOC116806870	LOC116806990
ATP5ME	DTWD2	HINT2	LOC115491151	NIPBL	REEP5	SNX18	UBQLN1	LOC116806871	LOC116806993
BDP1	ECPAS	HMGCR	LOC115491190	NLN	RFK	SNX2	UGCG	LOC116806872	SMAD4
BRIX1	ELAC1	HMGCS1	LOC115491229	NMRK1	RFX3	SNX24	UHRF2	LOC116806874	
BTF3	ELAVL2	HNRNPK	LOC115491243	NOL6	RGMB	SNX30	UTP15	LOC116806876	
C9orf72	ELP1	HSD17B4	LOC115491251	NSA2	RGP1	SPIN1	VCP	LOC116806879	
CAAP1	EMC4	HSDL2	LOC115491282	NTRK2	RGS7BP	SPTLC1	WDR36	LOC116806881	
CARNMT1	ERBIN	IDUA	LOC115491283	NUDT12	RHOBTB3	SREK1	WDR41	LOC116806882	
CBWD1	ERCC6L2	IL6ST	LOC115491288	NUP155	RIC1	SREK1IP1	WDR70	LOC116806883	
CCDC112	ERCC8	INIP	LOC116806725	PAIP1	RICTOR	SRFBP1	XPA	LOC116806886	
CCDC171	ERMP1	IPO11	LOC116806757	PAM	RIOK2	SRP19	ZCCHC9	LOC116806887	
CCNH	EXOSC5	ITGA2	LOC116806794	PARP8	RIT2	ST8SIA5	ZDHHC21	LOC116806888	
CDC37L1	FAM172A	KATNAL2	LOC116806830	PCGF3	RNF170	STARD4	ZFAND5	LOC116806889	
CDK7	FAM174A	KDM4C	LOC116806836	PELO	RNF180	STOML2	ZFR	LOC116806892	
CEMIP2	FAM219A	KIAA0825	LSM5	PGM5	RNF20	SUB1	ZFYVE16	LOC116806894	
CENPH	FANCG	KIAA1328	LYRM7	PHAX	RNF38	SYT4	ZNF131	LOC116806898	
CENPK	FBXL17	KIF2A	LYSMD3	PIAS2	RPL17	TARS1	ZSWIM6	LOC116806902	
CEP78	FBXO4	LMBRD2	MACIR	PIGG	RPL37	TBCA		LOC116806904	
CERT1	FCHO2	LNPEP	MAP3K1	PIGO	RPP25L	TENT2		LOC116806905	
CETN3	FEM1C	LOC100190031	MARVELD2	PIK3C3	RPS23	THAP1		LOC116806906	
CHD1	FER	LOC100217858	MBLAC2	PIP5K1B	RPS6	TMED7		LOC116806909	
CLTA	FKTN	LOC100219485	MCCC2	PJA2	SCAMP1	TMEM161B		LOC116806910	
COMMD10	FNTA	LOC100221697	MCTP1	PLAA	SECISBP2	TMEM167A		LOC116806911	

Table S2- Sex chromosomes consistently repressed or expressed across regions. Lists Z and W chromosome genes from Fig. 3e-f which exhibited sexually dimorphic expression in vehicle-treated, non-vocal learning related samples across brain regions.