Cellular profiling of a recently-evolved social behavior

Zachary V. Johnson*, Brianna E. Hegarty*, George W. Gruenhagen*, Tucker J. Lancaster, Patrick T. McGrath†, Jeffrey T. Streelman†

* these authors contributed equally to this work
† corresponding authors

ABSTRACT

Social behaviors are essential for survival and reproduction and vary strongly among individuals, species, and heritable brain diseases. The molecular and cellular bases of this variation are poorly resolved, and discovering them is necessary to understand how neural circuit and behavioral functions—and dysfunctions—vary in social contexts. Here we integrate single nucleus RNA-sequencing (snRNA-seq) with comparative genomics and automated behavior analysis to investigate the neurobiology of castle-building, a recently-evolved social, spatial, goal-directed, and repetitive construction behavior in Lake Malawi cichlid fishes. We simultaneously control for and analyze two biological variables correlated with castle-building behavior: quivering, a courtship “dance” behavior, and relative gonadal mass. We find signatures of building-, quivering-, and gonadal-associated neuronal excitation, gene expression, and neurogenesis in distinct cell populations. Converging lines of evidence support the involvement of estrogen, TrkB, and CCK signaling systems, and specific pallial excitatory neuronal subpopulations, in castle-building behavior. We show additional evidence that castle-building has evolved in part through genomic divergence in a gene module that is selectively expressed in stem-like quiescent radial glial cells (RGCs) lining the ventricular zone of the pallium. This RGC subpopulation exhibits signatures of a building-associated departure from quiescence, which in turn is associated with neuronal rebalancing in the putative fish homologue of the hippocampus. Our work supports an unexpected role for glia and neurogenesis in the evolution of social behavior, and more broadly shows how snRNA-seq can be used to systematically profile the cellular bases of previously unstudied social behaviors in new species systems.

INTRODUCTION

Social behaviors evolve rapidly and vary among individuals, species, and human brain disease states (Johnson and Young 2017). Different experimental traditions spanning genomics (C. R. Smith et al. 2008; Küpper et al. 2016; Lamichhaney et al. 2016; Bendesky et al. 2017; York et al. 2018; Pfenning et al. 2014; Dias and Walsh 2020; Stein et al. 2017), endocrinology (S. A. Juntti et al. 2016; Boender and Young 2020; Adkins-Regan 2013; O'Connell, Matthews, and Hofmann 2012; S. Ogawa et al. 2000; Heinrichs and Gaab 2007; Schiller, Meltzer-Brody, and Rubinow 2015), and circuit neuroscience (Gutzeit et al. 2020; Hung et al. 2017; Amadei et al. 2017; Anderson 2016; Gangopadhyay et al. 2021; Kohl et al. 2018; S. B. Nelson and Valakh 2015; Bachevalier and Loveland 2006) have advanced our understanding of the biological underpinnings of social behavior. Bridging these levels is challenging, and most efforts have focused on a relatively small number of model species—typically genetically inbred strains (S. Juntti 2019; Gallant and O'Connell 2020; Laurent 2020; Keifer and Summers 2016; Brenowitz and Zakon 2015; Jourjine and Hoekstra 2021; Johnson and Young 2018). Two drawbacks to this strategy are a lack of access to many social behaviors that are not expressed in traditional models, and a limited ability to understand how natural genetic variation generates variation in circuit and behavioral functions—and dysfunctions—in social contexts. The ability to profile many cell populations simultaneously in any species with a suitable reference genome erodes historical constraints that have faced investigations of non-traditional social behaviors and species systems. Here we integrate single nucleus RNA-sequencing (snRNA-seq) with automated behavior analysis and comparative genomics to investigate the molecular, genetic, and cellular substrates of castle-building, a recently-evolved (<1 Mya) social, goal-directed, repetitive and reproductive behavior in Lake Malawi cichlid (Cichlidae) fishes.

Cichlids are teleost (*Teleostei*) fishes, a group representing 40-50% of all living vertebrate species (Salzburger 2018). At teleosts, cichlids possess predicted homologues for ~80% of human disease-associated genes (Howe et al. 2013), and their brains are composed of conserved cell populations with conserved molecular, electrophysiological, morphological, transcriptional, and behavioral properties (O'Connell and Hofmann 2011; Xie and Dorsky 2017; Elliott et al. 2017; Jurisch-Yaksi, Yaksi, and Kizil 2020). For example, the teleost telencephalon contains conserved cell populations that regulate social behaviors across vertebrate lineages (O'Connell and Hofmann 2011). In Lake Malawi, ~800 cichlid species are behaviorally diverse but genetically similar, representing a powerful system for investigating the neurogenetic basis of behavioral variation (York et al. 2015; Baran and Streelman 2020; Johnson and Young 2018; Johnson, Arrojwala, et al. 2020; York et al. 2018; Ribbink et al. 1983; Johnson, Moore, et al. 2020). Approximately 200 species express bower construction behaviors during the breeding season, whereby males manipulate sand into species-specific structures for courtship (York et al. 2015; Johnson, Arrojwala, et al. 2020; Long et al. 2020). Many species dig crater-like "pits" while others build volcano-like "castles," and these behavioral differences are associated with genomic divergence in a ~19 Mbp chromosomal region enriched for genes that exhibit *cis*-regulated behavior-associated expression (York et al. 2018).

In this paper we investigate castle-building behavior in *Mchenga conophoros*, a Lake Malawi cichlid and an uncharted species in behavioral neuroscience. We map the cellular diversity of the telencephalon, identify cell type-specific signatures of active castle-building behavior, and link genomic divergence associated with castle-building evolution to specific cell populations. Our work shows how snRNA-seq profiling can be used to systematically identify converging lines of evidence for candidate genes, molecular systems, cell populations, and brain regions underlying new behaviors in new species systems.

RESULTS

Castle-building is associated with increased quivering behavior and gonadal physiology

We used an automated behavior analysis system (Johnson, Arrojwala, et al. 2020; Long et al. 2020) to monitor reproductive adult *Mchenga conophoros* males as they freely interacted with four reproductive adult females and sand over multiple days (Fig. 1A). Briefly, this system uses depth sensing to measure structural changes across the sand surface and action recognition to predict building and quivering behaviors from video data. We sampled pairs of males at the same time in which one male was actively castle-building (n=19) and the other was not ("control", n=19; pairs correspond to rows in Fig. 1B). For each subject, we also recorded the gonadal somatic index (GSI), a measure of relative gonadal mass that is correlated with

gonadal steroid hormone levels and social behaviors in cichlids (Maruska and Fernald 2010; Ramallo et al. 2015; Alward et al. 2019) (Table S1). The volume of sand displaced by males was strongly and positively correlated with the total number of building events predicted from video data by action recognition ($p=8.15x10^{-13}$, t-test). For simplicity, we used both depth and action recognition data to generate a single "Bower Activity Index" (BAI), which also differed strongly between building and control males (Fig. 1G; $p=4.24x10^{-8}$, paired t-test). Building males quivered more (a stereotyped courtship "dance" behavior; Fig. 1H; $p=9.18x10^{-6}$, paired t-test), and had greater GSIs (Fig. 1I; p=0.0142, paired t-test) than controls. Taken together, these results are consistent with castle-building, like many social behaviors in nature, being embedded within a suite of behavioral and physiological changes tied to reproduction.

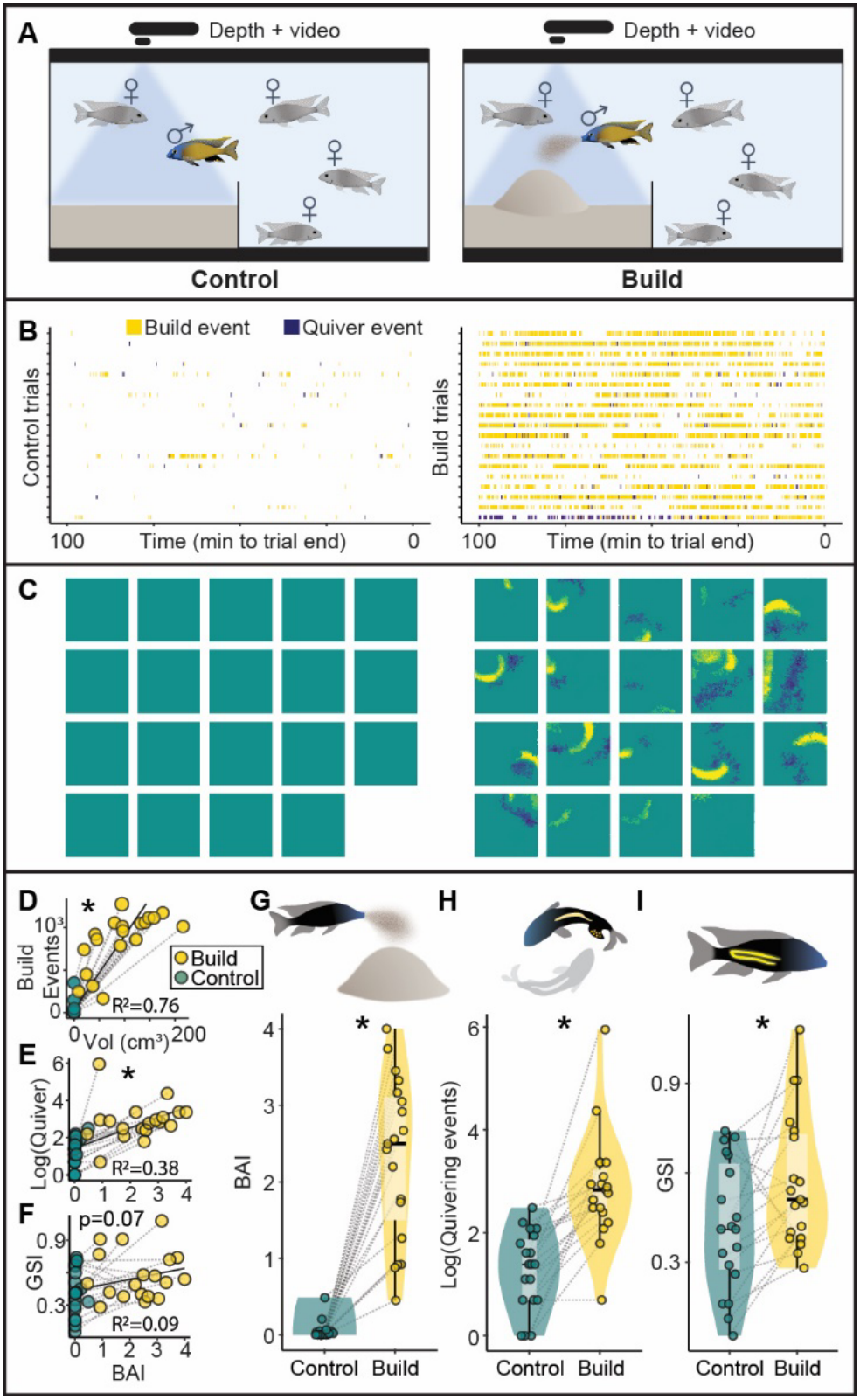

Figure 1. Castle-building is associated with increased quivering and relative gonadal mass. (A) 19 pairs of building (right) and control (left) males were sampled. Action recognition (B, yellow=building, blue=quivering, each trial is represented by a row, with pairs matched by row between left and right panels) and depth sensing (C, yellow=elevations, blue=depressions, each square represents total depth change for one trial, with pairs matched by row and column between left and right panels) revealed behavioral differences between building and control males. (D) Structural change measured through depth sensing (adjusted for body size) was strongly and positively correlated with building behaviors predicted through action recognition, and these measures were combined into a single Bower Activity Index (BAI, x-axis in E

and F). (E) BAI was positively correlated with quivering behaviors across trials, and (F) trended toward a positive correlation with GSI across trials. Building was associated with significantly greater BAI (G), quivering (H, log-normalized), and GSI (I). Gray lines in panels D-I show pairs of control and building males.

Telencephalic nuclei reflect major neuronal and non-neuronal cell classes

Telencephala (n=38) were combined into ten pools (n=5 behave, n=5 control, 3-4 telencephala/pool) for snRNA-seq (Fig. 2A). In total, >3 billion RNA reads were sequenced and mapped to the Lake Malawi cichlid *Maylandia zebra* reference genome (Conte et al. 2019). 33,674 nuclei (~900 nuclei/subject) passed quality control filters and were linked back to test subjects using genomic DNA. Coarse-grained clustering grouped nuclei into 15 "primary" (1°) clusters and finer-grained clustering grouped nuclei into 53 "secondary" (2°) clusters (ranging from 57-1,905 nuclei, Fig. 2B). Established marker genes revealed known neuronal and non-neuronal cell types (Fig. 2C), including excitatory (*slc17a6*+) and inhibitory (*gad2*+) neurons, oligodendrocytes and oligodendrocyte precursor cells (OPCs, *olig2*+), radial glial cells (RGCs, *fabp7*+), microglia, pericytes, and hematopoietic stem cells (Table S2). Additional analyses revealed nearly cluster-exclusive expression of many genes (Fig. 2D), and cluster-specific expression of genes encoding transcription factors (TFs; Fig. 2E-F) and neuromodulatory signaling molecules (Fig. 2G-H) that exhibit conserved neuroanatomical expression patterns in teleosts (Table S2). The relative proportions of clusters were consistent across individuals (Fig. 2I, see Table S3 for detailed 1° and 2° cluster information). For clarity, we assigned each 1° cluster a numeric identifier (1-15) followed by a label indicating one or more of these cell classes (e.g. for RGCs, "_RGC"). 2° cluster labels were rooted in these 1° labels, but with a second numeric identifier indicating the relative size within the corresponding "parent" 1° cluster (e.g. "4_GABA" is a 1° cluster expressing inhibitory neuronal markers, and "4.3_GABA" is the third largest 2° cluster within 4_GABA). Marker genes for every individual primary and secondary cluster were independently enriched ($q<0.05$) for eight GO categories related to cell morphology, connectivity, conductance, and signal transduction (Table S4), suggesting these were major axes distinguishing clusters in this study.

Clusters are more strongly distinguished by transcription factors and ligands than receptors

Previous work suggests that different categories of genes (e.g. TFs, ligands, receptors) may differ in transcriptional plasticity across cell populations (Moffitt et al. 2018; O'Connell and Hofmann 2012). We curated lists of genes encoding either 1) neurodevelopment/neuroanatomy-associated TFs (nTFs, n=43; Table S5) with conserved brain region-specific expression patterns in multiple teleost lineages (Table S2), 2) neuromodulatory ligands or related transporter proteins ("ligands", n=35), or 3) neuromodulatory receptors ("receptors", n=108), and investigated their expression patterns across clusters . 1° and 2° marker genes were more strongly enriched for nTFs and ligands compared to receptors (receptors versus nTFs, $p \leq 8.33 \times 10^{-4}$ for both 1° and 2° clusters, FET; ligands, $p \leq 0.0068$ for both; versus ligands, $p \geq 0.75$ for both, Fig. 2J), consistent with recent scRNA-seq analyses of the mouse hypothalamus (Moffitt et al. 2018). Notably, several nTFs involved in dorsal-ventral (DV) patterning in early neural development exhibited striking polarity in expression across clusters (Fig. 2F). For example, *dlx* genes and *isl1* mark the ventral telencephalon while *emx* genes mark the dorsal telencephalon during the neurula stage (Sylvester et al. 2013), suggesting that neuronal populations in adulthood may retain transcriptional signatures of their developmental origins. Together these data may reflect organizing principles whereby cell populations are more strongly aligned with transcriptional programs underlying neurodevelopment and ligand synthesis, and neuromodulatory receptors are expressed more promiscuously across cell populations.

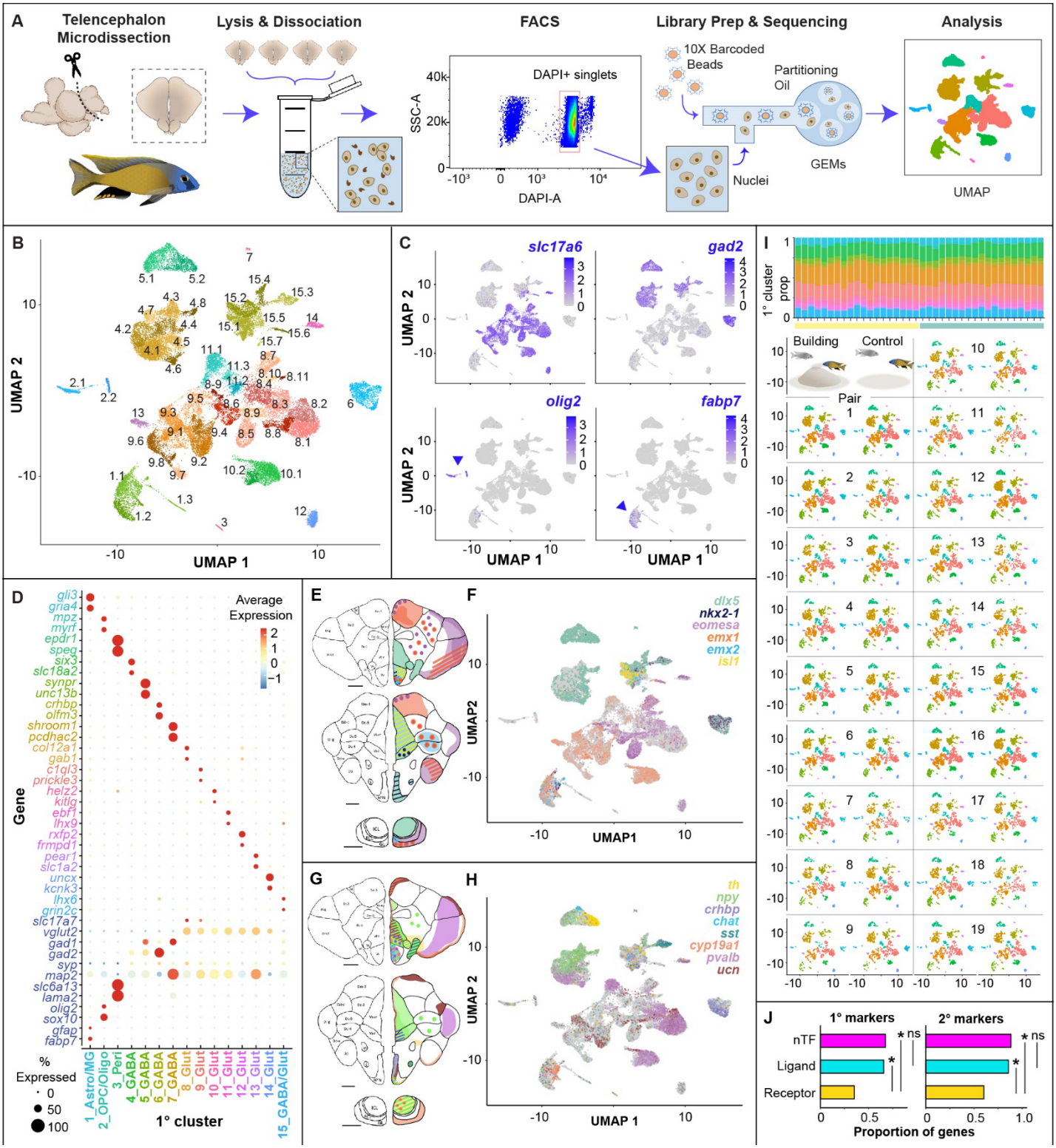

Figure 2. Molecular and cellular diversity of the cichlid telencephalon. (A) Schematic of experimental pipeline for snRNA-seq. (B) Nuclei cluster into 1° (n=15) and 2° (n=53) clusters. (C) Known marker genes reveal distinct clusters of excitatory neurons (*slc17a6*+), inhibitory neurons (*gad2*+), oligodendrocytes and oligodendrocyte precursor cells (*olig2*+), radial glial cells (*fabp7*+), as well as other less abundant cell types (not shown, see Table S2). (D) Cluster proportions are consistent across 38 males. (E) Clusters are distinguished by expression of cluster-exclusive and known cell type marker genes. Conserved nTFs (F, G) and ligands (and related genes; H, I) exhibit conserved neuroanatomical expression profiles in the teleost telencephalon and are distinctly expressed in specific clusters. (J) Genes encoding neuromodulatory receptors are more promiscuously expressed across 1° and 2° clusters than genes encoding neuromodulatory ligands or nTFs.

Building, quivering, and gonadal physiology are associated with transcriptional signatures of neuronal excitation in distinct cell populations

To identify candidate cell populations that may regulate castle-building behavior, we first investigated transcriptional signatures of neuronal excitation. Neuronal excitation triggers intracellular molecular cascades that induce transcription of conserved immediate early genes (IEGs) (Lyons and West 2011), and mapping IEG expression is a strategy for identifying neuronal populations that are excited by specific stimuli or behavioral contexts (Guzowski et al. 2005). However, IEG transcripts tend to be recovered at lower levels compared to other genes in sc/snRNA-seq data (Y. E. Wu et al. 2017; Lacar et al. 2016; Moffitt et al. 2018). To better track IEG responses, we identified genes that were selectively co-transcribed with three established IEGs (*c-fos*, *egr1*, *npas4*) independently across 2° clusters. In total, we identified 25 "IEG-like" genes (Table S6), most (17/25, 68%) of which had previously been identified as IEGs, but eight of which have not (to our knowledge, predicted homologues of human *DNAJB5, ADGRB1, GPR12, ITM2C, IRS2, RTN4RL2, RRAD*; Fig. 3A). These genes may include new conserved markers of neuronal excitation.

We assigned each nucleus an "IEG score," equal to the number of unique IEG-like genes expressed. To disentangle building-, quivering-, and GSI-associated signals, we tested a sequence of models in which these variables competed in different combinations to explain variance in IEG score. Effects were considered significant if the raw p-value was significant ($p<0.05$) in every model and if the FDR-adjusted harmonic mean p-value (hmp_{adj}) was significant across models ($hmp_{adj}<0.05$) (Wilson 2019). Building was associated with increased IEG expression in 9_Glut ($hmp_{adj}=0.0016$; Fig. 3B), a cluster distinguished by markers of the dorsal pallium (Martinelli et al. 2016). We expected that some behaviorally-relevant populations may not align with clusters. For example, neuropeptides can diffuse to modulate distributed cell populations expressing their target receptors (Johnson and Young 2017), and other behaviorally–relevant populations may represent a small proportion of one cluster. We therefore analyzed populations defined by nTF, ligand, and receptor genes, both within clusters and regardless of cluster. IEG score was associated with building, quivering, and GSI in distinct cell populations (Fig. 3B). Notably, building was associated with IEG score in 9_Glut, a cluster distinguished by markers of the dorsal and lateral pallium, in three populations defined regardless of cluster (*elavl4*+, *cckbr*+, *ntrk2*+), and in 4_GABA *htr1d*+, 4_GABA *vipr2*+, 15_GABA/Glut *tacr2*+, 11_Glut *cckbr*+, and 11.1_Glut *npr2*+ nuclei (Fig. 3C), consistent with a role for these molecular systems in the neural coordination of building. Quivering was associated with IEG score in 5.2_GABA *elav4*+ nuclei, a subpopulation strongly and/or selectively expressing a suite of genes expressed in the olfactory bulb granule cell layer and in newborn dopaminergic neurons in teleosts (e.g. *tac1*, *pax6*, *trh*, *th*, *dat*, *vmat*). These data are consistent with previous work showing activation of olfactory and dopaminergic circuitry during courtship in diverse systems (Keleman et al. 2012; van Furth, Wolterink, and van Ree 1995; Ishii and Touhara 2019; Louilot et al. 1991; Johnson and Young 2015). Behavior-associated IEG expression was most strongly associated with behavior expressed approximately 60 minutes prior to sample collection, consistent with previously reported nuclear IEG RNA time courses (Fig. 3D) (Lacar et al. 2016). Previous work has demonstrated behavior-associated shifts in correlated IEG expression across brain regions in diverse species, such that social context is associated with a shift in how strongly IEG expression in one region predicts IEG expression in another region across individuals (Johnson and Young 2017; Johnson et al. 2016, 2017; Hoke, Ryan, and Wilczynski 2005; Yang and Wilczynski 2007; Almeida et al. 2019). We found that populations exhibiting building-associated IEG expression also differed in correlated IEG expression across building versus control males (Fig. 3E), consistent with shifts in correlated excitation across cell populations during building. Because 1) the excitatory and inhibitory phenotypes and 2) the directionality of changes in correlated IEG expression were both known, simplified but testable circuit hypotheses about synaptic connectivity among populations can be generated. For example, because building is associated with a more negatively correlated Fos expression between 4_GABA *vipr2*+ nuclei (inhibitory) and other populations. A simple and testable model is that 4_GABA *vipr2*+ directly synapses onto these populations. More broadly, these results identify specific cell populations (e.g. 9_Glut) and molecular systems (e.g. TrkB, CCK) as candidate regulators of castle-building.

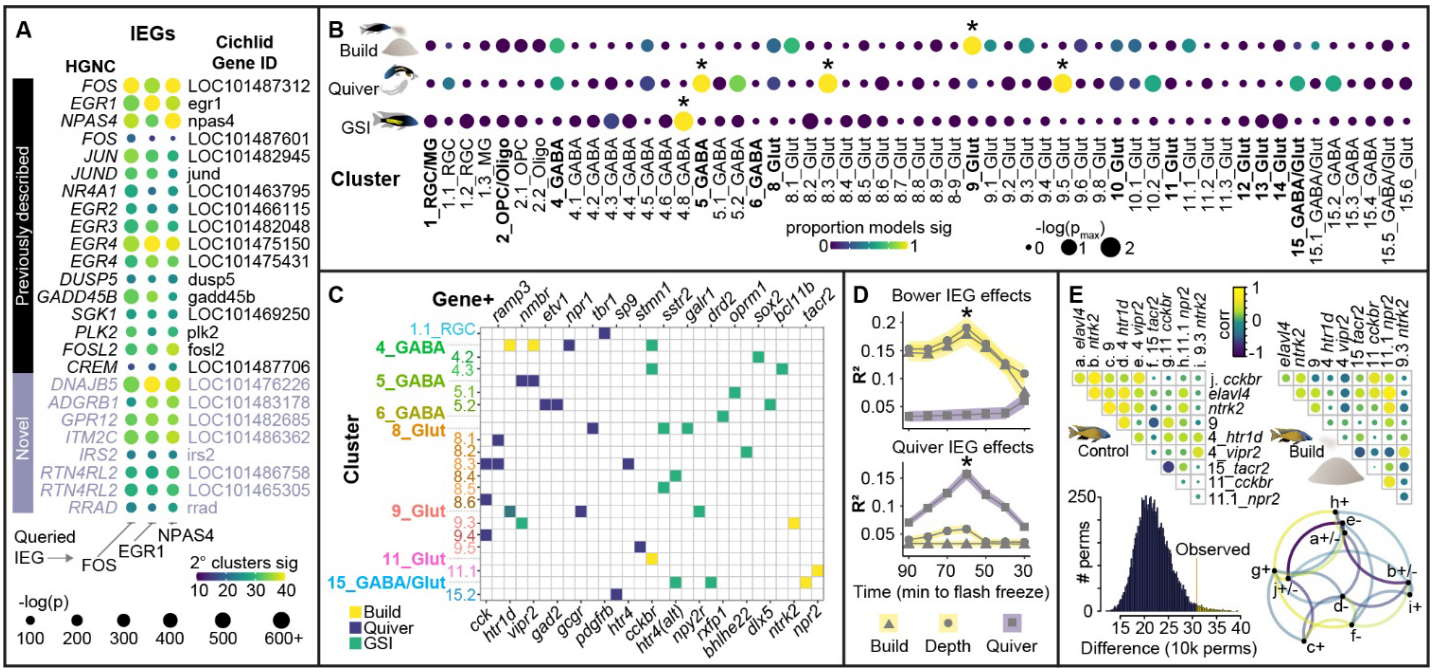

Figure 3. Distinct cell populations exhibit building-, quivering-, and gonadal-associated IEG expression. (A) 25 genes were selectively co-expressed with *c-fos*, *egr1*, and *npas4* across cell populations. Building-, quivering-, and gonadal-associated IEG expression was observed in distinct clusters (B) and gene-defined populations (C, filled squares indicate significant effects, q<0.05). (D) IEG expression was most strongly associated with the amount of building (top) and quivering (bottom) behavior performed approximately 60 minutes prior to tissue freezing. (E) Among populations exhibiting building-associated IEG expression, correlated IEG expression across individuals differed between control (left) and building (right) males (permuted null distribution shown in the bottom left; schematic indicating the directionality of the strongest changes, with excitatory and inhibitory phenotypes indicated by "+" and "-", respectively, shown in bottom right; letter codes for cell populations match letter codes above heatmaps).

A minority of neuronal populations account for the majority of building-associated gene expression

Changes in social behavior are associated with changes in brain gene expression in diverse lineages (Robinson, Fernald, and Clayton 2008; Baran and Streelman 2020; Patil et al. 2021; York et al. 2018), but the underlying cell populations driving these effects are usually not determined. We performed an unsupervised analysis to identify differentially expressed genes (DEGs) in specific clusters. A relatively small subset of neuronal clusters accounted for a disproportionate number of building-associated DEGs (bDEGs), a pattern that was also true of quivering-associated DEGs (qDEGs) and gonadal-associated DEGs (gDEGs; Fig. 4A, dot color and size reflect number and proportion of DEGs across clusters, respectively). bDEGs were overrepresented in three excitatory neuronal clusters (8_Glut, 9_Glut, 10_Glut; $q \leq 1.83 \times 10^{-4}$ for all), qDEGs were overrepresented in two neuronal clusters (15_GABA/Glut, 11_Glut, q≤0.036 for both), and gDEGs were overrepresented in one inhibitory neuronal cluster (5_GABA, $q = 1.30 \times 10^{-5}$). bDEGs were overrepresented in a suite of aligned 2° clusters ($q \leq 6.69 \times 10^{-4}$ for all), qDEGs were overrepresented in 15.2_GABA, 8.1_Glut, and 8.6_Glut (q≤0.0074 for all), and gDEGs were overrepresented in 8.3_Glut and 8.4_Glut (q≤0.039 for both). Thus, distinct clusters were overrepresented for bDEGs, qDEGs, and gDEGs. Interestingly, a common set of bDEGs, gDEGs, and qDEGs were the same individual genes (n=81), consistent with behavior and gonadal hormones recruiting similar transcriptional programs in distinct populations (Fig. 4B). These results highlight the potential involvement of a small set of 2° neuronal clusters (8.3_Glut, 8.9_Glut, 9.7_Glut, 10.1_Glut) in castle-building behavior.

Behavior-associated DEGs exhibited a strong direction bias, and were predominantly upregulated in both 1° and 2° clusters ($p \leq 1.39 \times 10^{-12}$ for all, Fig. 4C). In contrast, gDEGs tended more modestly toward upregulation in 1° clusters (1° gDEG effects, $p = 2.09 \times 10^{-5}$) and were not directionally biased in 2° clusters (2° gDEG effects, p=0.92). Upregulated bDEGs, qDEGs, and gDEGs were each independently enriched for a large number of the same GO terms (q<0.05 for 499 GO Biological Processes, 147 GO Cellular Components, and

111 GO Molecular Functions), the strongest of which were related to synaptic transmission and plasticity (e.g. "synaptic signaling," $q \leq 3.54 \times 10^{-50}$ for all; "regulation of synaptic plasticity," $q \leq 1.83 \times 10^{-18}$ for all) or cell differentiation and neurogenesis (e.g. "nervous system development," $q \leq 6.93 \times 10^{-47}$ for all; "neurogenesis," $q \leq 4.49 \times 10^{-35}$ for all; "cell morphogenesis involved in neuron differentiation," $q < 6.96 \times 10^{-29}$ for all; Fig. 4D), consistent with behavior- and gonadal-associated regulation of synaptic function and cell morphogenesis.

Estrogen has been linked to both neuronal excitability and neurogenesis and regulates social behavior in diverse species (Diotel et al. 2013; Duarte-Guterman et al. 2015; Kelly and Rønnekleiv 2009; Sarkar et al. 2008). Estrogen can impact gene expression by binding to estrogen receptors (ERs), forming a complex that is translocated into the nucleus and acts as a TF by binding to Estrogen Response Elements (EREs) in DNA (Klinge 2001; Amenyogbe et al. 2020). bDEGs, gDEGs, and qDEGs were independently enriched for EREs, consistent with a role for estrogen in modulating behavior- and gonadal-associated gene expression ($p \leq 2.92 \times 10^{-4}$ for all; Fig. 4E; ERE-containing gene list in Table S7). ERE-containing bDEGs (n=22 unique genes) were most strongly enriched for GO terms including "modulation of chemical synaptic transmission" (top GO Biological Process, $q = 2.30 \times 10^{-4}$) and "Schaffer collateral - CA1 synapse" (top Cellular Component, $q = 2.22 \times 10^{-5}$), highlighting building-associated estrogenic modulation of synaptic function as a potential player in castle-building.

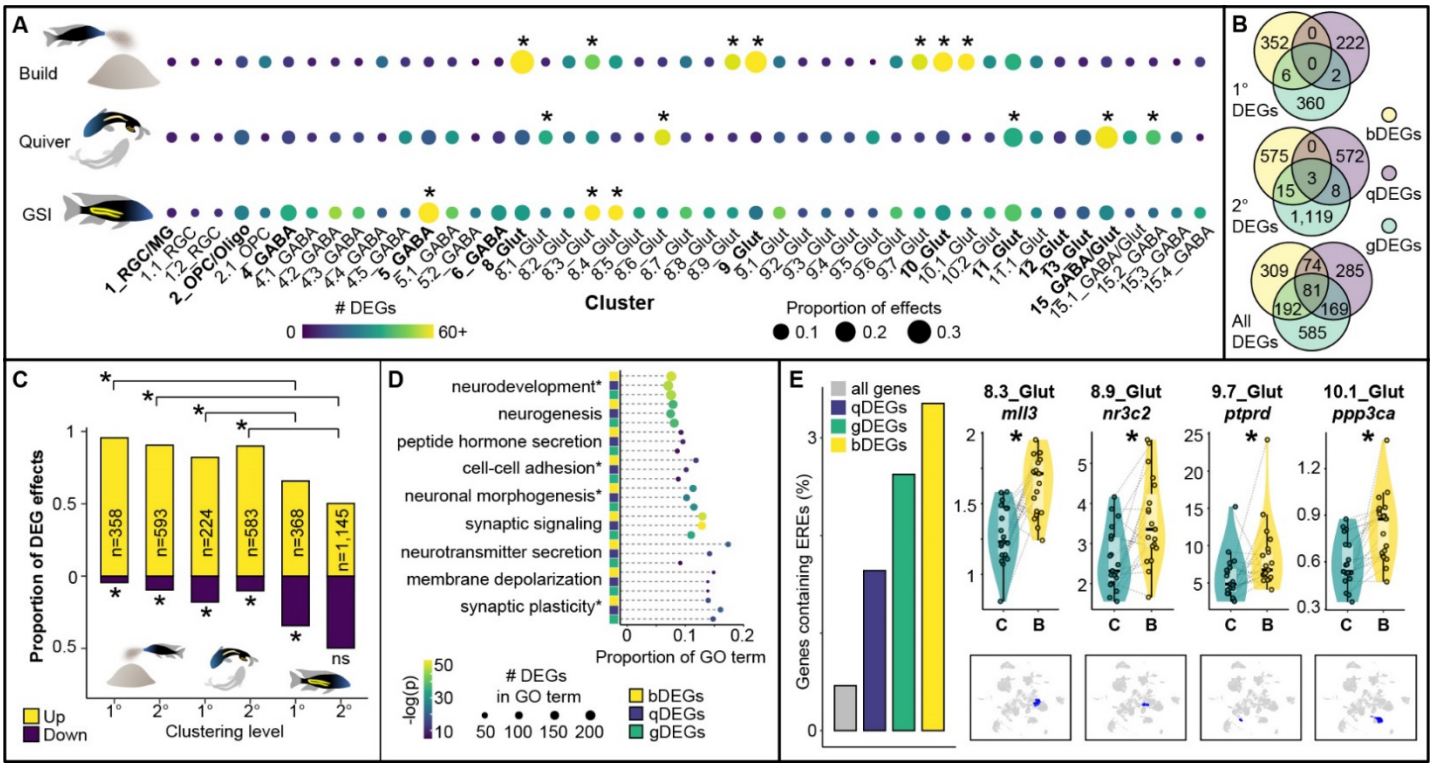

Figure 4. Building, quivering, and GSI are associated with distinct patterns of cell type-specific gene expression. (A) Distinct 1° and 2° clusters show a disproportionate number of bDEGs, qDEGs, and gDEGs. (B) A set of 81 genes exhibits building-, quivering, and gonadal-associated expression in largely non-overlapping clusters. (C) Behavior-associated gene expression is driven by upregulation, whereas gonadal-associated gene expression is driven by a balance of up- and downregulation. (D) bDEGs, qDEGs, and gDEGs are enriched for GO terms related to synaptic structure, function, and plasticity; neurotransmission; and neurogenesis. (E) bDEGs, qDEGs, and gDEGs are enriched for EREs. Violin plots show cluster-specific ERE-containing bDEG effects and feature plots below show the clusters (blue) in which each effect was observed. GO terms followed by asterisks are abbreviated.

Castle-building is associated with increased expression of genes that positively regulate neurogenesis

Most (6/7) clusters enriched for bDEGs did not exhibit building-associated IEG expression, consistent with a large proportion of building-associated gene expression being driven by processes other than neuronal

excitation. Based on the enrichment of neurogenesis-related GO terms among bDEGs, we reasoned that neurogenesis and related processes (e.g. axon growth, dendritic branching) may underlie a large proportion of building-associated gene expression and be important for castle-building behavior. To further investigate this, we identified 87 genes with the GO annotation “positive regulation of neurogenesis” in both zebrafish and mice (“proneurogenic” genes, pNGs, Table S8). Six 1° (8_Glut, 9_Glut, 10_Glut, 11_Glut, 15_GABA/Glut, 4_GABA) and ten aligned 2° neuronal clusters exhibited building-associated increases in pNG expression (hmp_{adj}≤0.020 for all; Fig. 5A), including 5/7 clusters that exhibited a disproportionate number of bDEGs. The most significant building-associated pNG expression was observed in 8_Glut (Fig. 5B, $hmp_{adj}=1.23x10^{-17}$), a cluster distinguished by markers of Dl, the putative hippocampal homologue in fish. In contrast to building, gonadal-associated pNG expression was increased in 10.2_Glut (hmp_{adj}=0.010) and decreased in 4.8_GABA (hmp_{adj}=0.0048), and quivering was not associated with pNG expression in any 1° or 2° cluster (Fig. 5A). Notably, the magnitude of effect (β) estimates for building-associated pNG expression in 2° clusters were always greater than in their “parent” 1° clusters, and many gene-defined subpopulations within clusters exhibited stronger building-associated pNG expression than their parent clusters. For example, within 15_GABA/Glut, building-associated pNG estimates were >3x greater in subpopulations defined by expression of *adra2b* ($β_{cond=}$=0.188) and *esr2* ($β_{cond}$=0.154) compared to 15_GABA/Glut as a whole ($β_{cond}$=0.048). Among 2° clusters, the most extreme cases included 8.2_Glut *drd4*+, 8.4_Glut *htr4*+, 9.1_Glut *sstr5*+, 9.6_Glut *htr4*+, 10.1_Glut *ntrk2*+ nuclei, and 11.1_Glut *ntrk2*+ nuclei (hmp_{adj}≤0.018 for all). Among populations defined regardless of cluster, those exhibiting building-associated pNG expression were disproportionately defined by neuromodulatory receptor and ligand genes versus nTFs (receptors versus nTFs, q=0.011; ligands versus nTFs, q=0.017; FET), and those exhibiting the strongest building-associated pNG expression (β) effects were disproportionately defined by neuromodulatory receptor genes (q=0.011; Fig. 5D), and by ERs in particular (q=0.034; Fig. 5E), consistent with a large body of literature supporting relationships between estrogen and neurogenesis (Diotel et al. 2013; Duarte-Guterman et al. 2015). These results highlight specific molecular systems (e.g. estrogen, serotonin, TrkB) as potential molecular systems involved in building-associated neurogenic changes.

Castle-building is associated with putative hippocampal neuronal rebalancing

During neurogenesis, new neurons differentiate into specific neuronal populations (Mira and Morante 2020; Götz and Huttner 2005). We reasoned that if building is associated with cell type-specific neurogenesis, then it may be associated with changes in the relative proportions of specific neuronal populations. Building was associated with decreased proportions of 8.1_Glut ($q=7.67x10^{-4}$; Fig. 5F,G) and increased proportions of 8.4_Glut (q=0.013; Fig. 5F,H). The relative proportions of 8.4_Glut and 8.1_Glut were negatively correlated across subjects, such that greater proportions of 8.4_Glut predicted lesser proportions of 8.1_Glut (R=-0.50, p=0.0012; Fig. 5I). We hypothesized that this rebalancing could be mediated in part by a neurogenesis-mediated influx of new neurons into 8.4_Glut. Indeed, 8.4_Glut was among the 2° clusters that exhibited building-associated increases in pNG expression, and a more direct test showed that pNG expression in 8.4_Glut was positively associated with its relative proportion (R=0.33, p=0.041). Together, these data are consistent with a building-associated rebalancing of two excitatory neuronal populations mediated in part by increased influx of new neurons into 8.4_Glut.

Building is associated with changes in RGC biology

RGCs are a primary source of new neurons in adult teleosts (Ganz and Brand 2016), and we therefore reasoned that signatures of neurogenesis may be downstream effects of changes in RGCs. We first investigated building-associated gene expression within RGCs (1.1_RGC and 1.2_RGC pooled). We identified 25 bDEGs that were collectively enriched for "neuron development” (top GO Biological Process, $q=8.18x10^{-4}$) it’s as well as “astrocytic glutamate-glutamine uptake and metabolism” (top Pathway, q=0.0010) and "synapse" (top GO Cellular Component, q=0.0015). RGC bDEGs included *cyp19a1* (encodes brain aromatase, upregulated; Fig. 5J). Aromatase converts testosterone to brain-derived estrogen and has been previously linked to RGC function and neurogenesis (Pellegrini et al. 2016).

RGCs can occupy distinct functional states including quiescence, cycling, and neuronal differentiation (Jurisch-Yaksi, Yaksi, and Kizil 2020; Adolf et al. 2006; Labusch et al. 2020). We re-clustered RGCs (independently of other nuclei) into 11 subclusters (RGC_0-RGC_{10}; Fig. 5K). We next assigned each nucleus a quiescence, cycling, and neuronal differentiation score based on established marker genes (Table S9), and

analyzed building-associated differences in these scores across subclusters. Building was associated with decreased quiescence score in RGC_2 (hmp_{adj}=0.010; Fig. 5L), but was not associated with quiescent, cycling, or neuronal differentiation score in any other subcluster. Analysis of building-associated gene expression across subclusters further revealed that 19/61 subcluster bDEGs were in RGC_2, and 18/19 effects reflected building-associated downregulation. The strongest enrichment hit for RGC_2 bDEGs was GO Cellular Component "postsynaptic Golgi apparatus" (q=0.0011), a possible reflection of building-associated changes in neuron-glia communication in this subpopulation. *cyp19a1* was excluded from analysis in several subclusters because it was not detected in all build-control pairs; however, a targeted analysis revealed that building-associated increases in *cyp19a1* were driven by RGC_3 (hmp_{adj}=0.018; Fig. 5M), a subpopulation distinguished by *lhx5* and *gli3*, both nTFs that regulate neurogenesis in mammals (Zhao et al. 1999; Hasenpusch-Theil et al. 2018). Lastly, because RGC subclusters strongly aligned with RGC functional states, we reasoned that building-associated changes in RGC function may also manifest as building-associated changes in RGC subcluster proportions. Indeed, building was associated with an increase in the relative proportion of RGC_4 (q=0.0017; Fig. 5N), a subcluster positioned between nuclei expressing markers of quiescence and nuclei expressing markers of cycling. These data support building-associated changes in RGC biology, and highlight RGC_2, RGC_3, and RGC_4 as candidate players in building-associated and RGC-mediated neurogenesis.

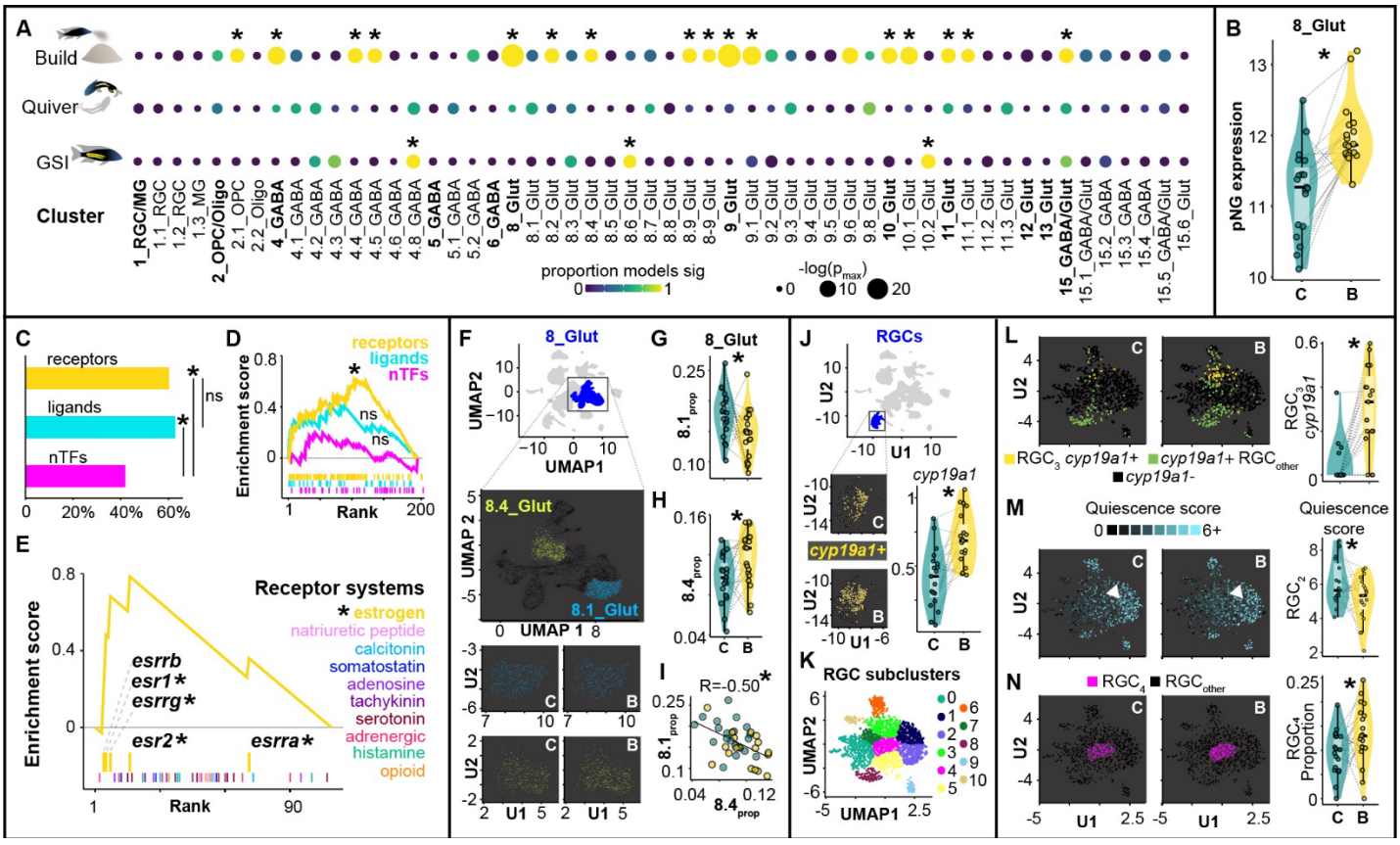

Figure 5. Behavior and gonadal physiology are associated with transcriptional signatures of neurogenesis in distinct cell populations. (A) Bower construction, but not quivering behavior, is associated with increased pNG expression in a large set of 1° and 2° clusters, whereas GSI is associated with increased and decreased pNG expression in just three 2° clusters. (B) The most significant building-associated pNG expression is observed in 8_Glut. (C) Gene-defined populations that exhibit building-associated pNG expression are disproportionately defined by genes encoding receptors and ligands. (D) The strongest building-associated pNG expression tends to occur in populations defined by neuromodulatory receptors, particularly in ER-expressing populations (E). (F-I) Building is associated with a shift in the relative proportions of two 2° neuronal clusters within 8_Glut. (J) RGCs exhibit building-associated *cyp19a1* expression. (K-N) reclustered RGC subpopulations exhibit building-associated *cyp19a1* expression (L), signatures of decreased quiescence (M), and increases in proportion (N).

Genes that have diverged in castle-building lineages are upregulated in reproductive contexts

Castle-building behavior has previously been linked to a ~19 Mbp region on Linkage Group 11 (LG11), within which genetic variants have diverged between closely-related castle-building and pit-digging lineages (York et al. 2018; Patil et al. 2021). Within this region, additional comparative genomics analyses revealed 165 "castle-divergent" genes (CDGs) that also showed signatures of divergence between castle-building lineages and more distantly-related species that do not build bowers (Fig. 6A; Table S10). CDGs were enriched for "carboxylic acid binding," "small conductance calcium-activated potassium channel activity," "ionotropic glutamate receptor signaling pathway," "proximal/distal pattern formation," "cell-cell junction," and human brain disease-associated cytobands (q≤0.047 for all, Table S5). CDGs were expressed at higher levels in the telencephalon compared to neighboring genes in the same 19Mbp region (~2.9x greater expression, permutation test, $p=1.42x10^{-5}$) and compared to other genes throughout the genome (~2.6x greater expression, $p=1.77x10^{-6}$). CDGs were also overrepresented among 1° and 2° cluster markers (versus neighboring LG11 genes, $p\leq1.66x10^{-9}$ for both; versus all other genes, $p\leq1.43x10^{-11}$ for both, FET), and among upregulated bDEGs, qDEGs, and gDEGs (versus neighboring LG11 genes, p≤0.0044 for all; versus all other genes, p≤0.0066 for all, FET; Fig. 6B). Taken together, these data support the behavioral significance of CDGs in the telencephalon, and more broadly that castle-building evolution has targeted genes that are upregulated during reproductive contexts.

Castle-divergent genes are enriched in quiescent RGC subpopulations

CDGs were most strongly enriched in non-neuronal clusters 2.1_OPC, 1.1_RGC, and 1.2_RGC, followed by neuronal clusters 4.3_GABA and 4.4_GABA and gene-defined populations 5.2_GABA *th*+, and 9_Glut *hrh3*+ (Fig. 6C; Table S11). We hypothesized that co-upregulation of subsets of CDGs in the same nuclei may drive cluster-specific enrichment patterns. A WGCNA based analysis revealed a module of 12 CDGs that were more strongly co-expressed than other CDGs (stronger correlation coefficients, Welch t-test, $p=8.83x10^{-14}$; stronger silhouette widths, Welch t-test, p=0.016; Fig. 6D). Across clusters, the module was most strongly enriched in 1.2_RGC ($p_{perm}=0$, Cohen's d=4.22), and was less strongly enriched in 1.1_RGC (Cohen's d=2.86; Fig. 2E), suggesting differences in expression among RGC subpopulations (Table S12). Among reclustered RGCs, CDG module expression mirrored expression of genetic markers of RGC quiescence (Fig. 2F). Indeed, the CDG module was positively associated with markers of RGC quiescence (R=0.34, $p=3.21x10^{-52}$; $p_{perm}=0$; Fig. 6G); and was negatively associated with cycling score (R=-0.089, $p=9.90x10^{-5}$; $p_{perm}=0$) and neuronal differentiation score (R=-0.065, p=0.0048; $p_{perm}=0$). The module was also enriched in subclusters RGC_1 ($p_{perm}=0.0196$) and RGC_2 ($p_{perm}=0.046$; Fig. 6H; Table S12), both of which selectively expressed genetic markers of RGC quiescence. Lastly, analysis of co-expression between the module and known TFs (n=999) identified *npas3* as the most strongly co-expressed TF with the module (R=0.47, $q=3.19x10^{-100}$). *npas3* suppresses proliferation in human glioma, is strongly expressed in quiescent neural stem cells, and is downregulated during hippocampal neurogenesis in mice (Moreira et al. 2011; Shin et al. 2015). Together these data support that CDG module expression is positively related to RGC quiescence.

An RGC subpopulation links evolution, behavior, and neuronal rebalancing in the putative fish hippocampal homologue

We hypothesized that the CDG module may regulate building-associated changes in RGC function and neurogenesis. In support of this, building was associated with a decrease in CDG module score in RGC_2 ($hmp_{adj}=0.027$; Fig. 6I), and an increase in CDG module score in RGC_8 ($hmp_{adj}=0.010$). *wdr73* was the only individual CDG module gene that exhibited building-associated expression and was selectively downregulated in RGC_1 and RGC_2 ($hmp_{adj}\leq4.54x10^{-89}$ for both; RGC_2 effect in Fig. 6J). Taken together, these data support a building-associated downregulation of *wdr73* and the CDG module in RGC_2, consistent with a departure from quiescence in this population. We hypothesized that building-associated changes in RGC biology may be related to building-associated neuronal rebalancing. Lasso, elastic net, and ridge regularization as well as mediation analysis supported *wdr73* expression in RGC_2 as the top candidate mediator of the building-associated rebalancing between 8.4_Glut and 8.1_Glut (Fig. 6K). Consistent with this, the 8.4_Glut:8.1_Glut ratio was predicted by RGC_2 CDG module score ($R=-0.52$, $p=6.91x10^{-4}$; Fig. 6L), *wdr73* expression ($R=-0.62$, $p=3.31x10^{-5}$; Fig. 6M), quiescent score (R=-0.42, p=0.0094), and *npas3* expression ($R=-0.52$, $p=8.20x10^{-4}$). All of these relationships were observed within building males only (8.4_Glut:8.1_Glut ratio versus RGC_2 CDG module score, R=-0.51, p=0.024; quiescent score, R=-0.42,

p=0.059; versus RGC_2 *wdr73* expression, R=-0.59, p=0.0074; *npas3* expression R=-0.65, p=0.0027), but not within controls (p≥0.14 for all). In contrast, none of these relationships were evident in RGC_1, regardless of whether the analysis was conducted across all subjects (p≥0.13 for all) or restricted to building males (p≥0.074 for all). Anatomically distinct RGCs in the teleost telencephalon send processes into distinct brain regions (Fig. 6N). Spatial profiling revealed that RGC_2 was anatomically positioned along the pallial but not subpallial ventricular zone (Fig. 6O,P), consistent with a potential to supply neurons to dorsal pallial regions. 8.4_Glut and 8.1_Glut respectively mapped to ventral and dorsal Dl, the putative fish homologue of the mammalian hippocampus (Fig. 6O,P). Taken together, these data support a relationship between building-associated expression of the CDG module in RGC_2 and hippocampal-like neuronal rebalancing.

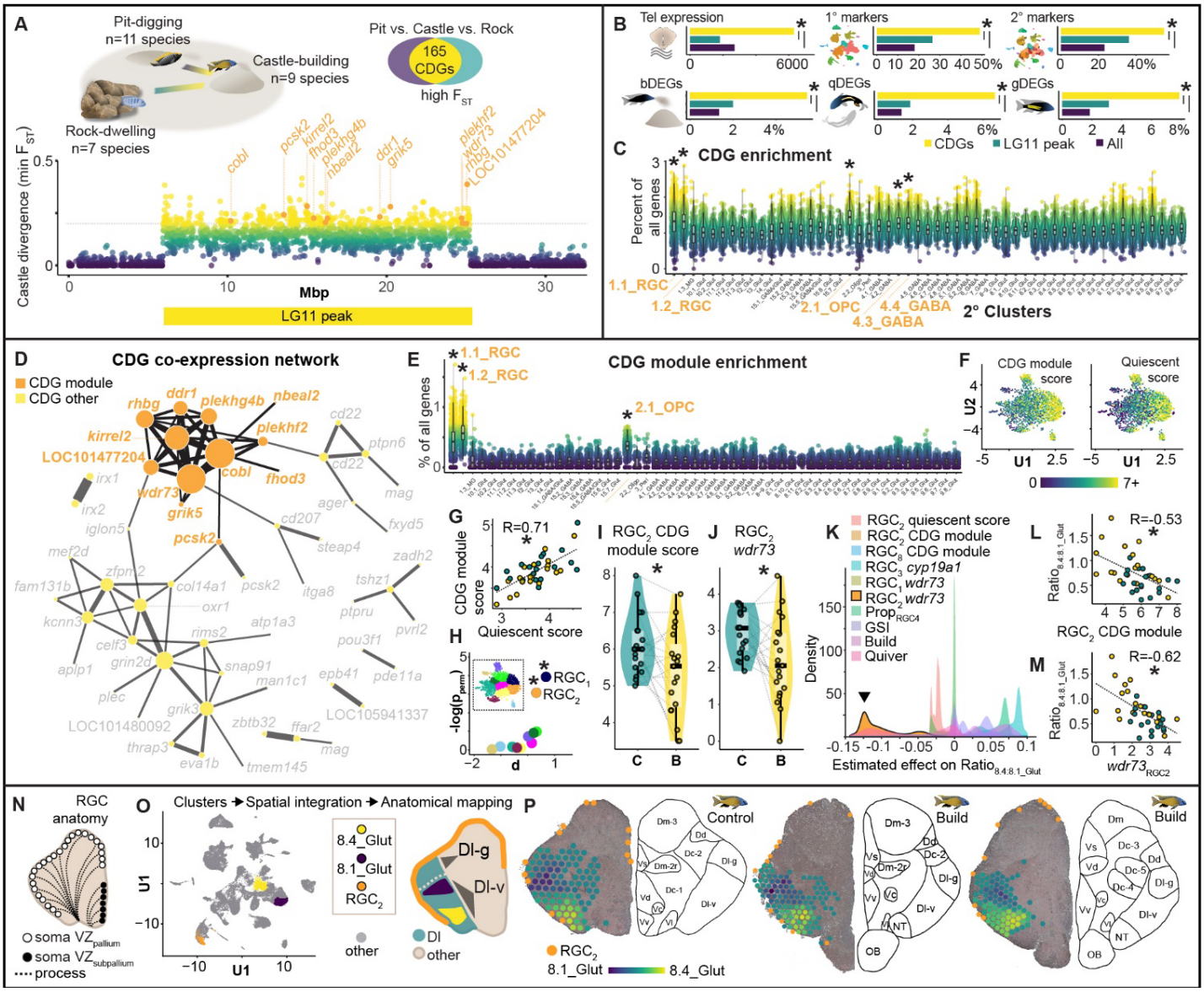

Figure 6. Genomic signatures of castle-building evolution link behavior, RGC function, and hippocampal-like neuronal rebalancing. (A) Comparative genomics identifies 165 CDGs. (B) CDGs are enriched in the telencephalon, among 1° and 2° cluster marker genes, and among bDEGs, qDEGs, and gDEGs. (C) CDGs are most strongly enriched in non-neuronal populations. (D) A CDG module (n=12 genes) is strongly co-expressed across nuclei. (E) The CDG module is most strongly enriched in radial glia. (F) CDG module expression across RGC subclusters mirrors expression of quiescent markers. (G) CDG module expression is correlated with expression of of RGC quiescence markers. (H) RGC_1 and RGC_2 are enriched for the CDG module. (I) RGC_2 exhibits building-associated decreases in expression of the CDG module and *wdr73* in particular (J). (K) *wdr73* expression in RGC_2 is the strongest predictor of neuronal rebalancing between 8.4_Glut and 8.1_Glut. (L-M) CDG module and *wdr73* expression are strongly associated with building-associated neuronal rebalancing between 8.4_Glut and 8.1_Glut. (N) RGCs differ in morphology,

function, and anatomical distribution (e.g. pallial versus subpallial ventricular zones). (O) Spatial profiling enables neuroanatomical mapping of RGC2, 8.1_Glut, and 8.4_Glut. (P) RGC2 (orange dots) aligned with pallial but not subpallial VZ, and 8.1_Glut versus 8.4_Glut aligned with dorsal versus ventral Dl-v, respectively.

DISCUSSION

The diversity of social behaviors in nature is an opportunity to discover organizing principles by which conserved biological systems generate variable neural and behavioral responses to social stimuli (Johnson and Young 2018). The ability to functionally profile many cell populations in under- and unstudied behavioral and species systems will be a boon to this endeavor. Here we integrated snRNA-seq with comparative genomics and automated behavioral analysis to systematically investigate the neurobiological substrates of castle-building in *Mchenga conophoros*. We first charted the cellular diversity of the telencephalon, and then functionally profiled behavior- and gonadal-associated gene expression as well as genomic signatures of behavioral evolution across clusters. Our analyses supported building-associated changes in cell type-specific neuronal excitation, gene expression, neurogenesis, and glial function. Multiple lines of evidence converged on estrogen, neurogenesis, 8_Glut and 9_Glut, and RGC-mediated hippocampal-like neuronal rebalancing as top candidate regulators of castle-building behavior.

Signatures of neuronal excitation reveal behavioral specificity and populations activated during building

Different social behaviors are regulated by distinct neural circuits and/or circuit activities in the brain (Newman 1999; Goodson 2005; Amadei et al. 2017; Kimchi, Xu, and Dulac 2007; Dulac, O'Connell, and Wu 2014). However, most tools cannot functionally profile many heterogeneous cell populations and simultaneously track their biological identities. Three studies in mice have supported the promise of sn/scRNA-seq technologies for mapping behavior-associated IEG expression (Lacar et al. 2016; Moffitt et al. 2018; Y. E. Wu et al. 2017); however, all three studies were conducted in the same genetically inbred C57BL6/J mouse strain, and thus cells could not be matched back to individuals after pooling. In our study, we leveraged natural genetic variation among individuals to trace ~34,000 nuclei back to 19 actively building and 19 control males. This allowed us to identify building-associated signals while simultaneously accounting for variance explained by other behavioral and biological factors that co-varied with building. Our analysis revealed novel IEG-like genes and distinct building-, quivering-, and GSI-associated signatures of neuronal excitation across cell populations. The only cluster exhibiting building-associated IEG expression was 9_Glut. 9_Glut was distinguished by genetic markers of Dd, a region that innervates Dl in a many-to-one fashion in other fish, mirroring the conserved "pattern separator" circuit organization of hippocampal cornu ammonis (CA) subregions and the dentate gyrus in mammals (Elliott et al. 2017). *ntrk2*+ nuclei also exhibited building-associated IEG expression, highlighting the TrkB system as a candidate player in castle-building. TrkB is a receptor that transduces activity-dependent signals into downstream modulation of neuronal differentiation, morphogenesis, survival, and long term potentiation (LTP) (Badurek et al. 2020; Lipsky and Marini 2007). Consistent with this, *ntrk2*+ nuclei also exhibited both building-associated pNG expression, suggesting this molecular pathway may link building-associated circuit activity to building-associated neuronal plasticity. The only other population that exhibited both building-associated IEG and pNG expression was defined by expression of *cckbr* (encodes Cholecystokinin B Receptor). Interesting, this receptor has recently been linked to NMDA receptor-mediated LTP and hippocampal neurogenesis in mice (Asrican et al. 2020; Chen et al. 2019).

A role for neurogenesis in the evolution and expression of social behavior

Multiple analyses supported neurogenesis as a top candidate and central player in castle-building behavior. Building was associated with upregulation of genes (bDEGs) related to neurogenesis as well as increases in pNG expression across clusters and gene-defined populations. Many pNGs regulate morphogenic processes integral to neuronal differentiation but that can also occur independently in mature neurons (e.g. axon growth, dendritic arborization, and/or dendritic spine formation). Thus, one potential explanation for these data is that building triggers cell type-specific forms of neuronal plasticity, similar to other forms of learning and behavior in rodents (Lai, Adler, and Gan 2018; Villanueva Espino, Silva Gómez, and Bravo Durán 2020; Hofer et al. 2009). Notably, building was associated with strong increases in pNG expression in both 8_Glut and 8.4_Glut

specifically, and with shifts in the relative proportions of 8.4_Glut and 8.1_Glut. These two populations mapped to different subregions of the putative fish hippocampal homologue, consistent with a growing body of literature demonstrating the importance of the hippocampus in social and spatial behaviors in mammals. Because neurogenesis is not a rapid process, it is difficult to interpret how neuronal rebalancing may be explained by behavior in recent hours. One possibility is that recent building in our paradigm was correlated with recent building spanning multiple previous days, and that these signals reflected longer-term and sustained changes in neurogenesis over the full course of building. Within this framework, it is intriguing to speculate that the rebalancing may be related to spatial representations of the bower structure and/or the territory, which need to be updated in the lab and in the wild. Lastly, building-associated neurogenesis was further supported by multiple signals in radial glia, including signals that we traced from genomic signatures of castle-building evolution, a topic we will return to later. Together our data suggest that programs underlying cell type-specific neurogenesis may have been integral to the evolution of castle-building behavior, and may also be important for its active expression. More broadly, these results fit with a large body of work showing changes in brain region-specific neurogenesis in other social and reproductive contexts across species (Walton, Pariser, and Nottebohm 2012; Bedos, Portillo, and Paredes 2018; Almli and Wilczynski 2012; Balthazart and Ball 2016; Maruska, Carpenter, and Fernald 2012; Dunlap, Chung, and Castellano 2013; Lévy et al. 2017).

Estrogenic substrates of male social behavior

Estrogen is a female gonadal steroid hormone that can be synthesized in the male brain via conversion of testosterone to estrogen by aromatase (L. R. Nelson and Bulun 2001). In the brain, estrogen can exert its effects at multiple levels, for example by regulating gene transcription (via EREs), neuronal excitability, synaptic plasticity, neurogenesis, and G-protein coupled receptor signaling (Kelly and Rønnekleiv 2009). Multiple lines of evidence supported a potential role for estrogen in the neural coordination of building. First, bDEGs (as well as qDEGs and gDEGs) contained canonical EREs, consistent with a role for estrogen in modulating building-associated gene transcription. Out of all GO terms, ERE-containing bDEGs were most strongly enriched for “Schaffer collateral - CA1 synapse” (driven by building-associated expression of *cacng2*, *ppp3ca*, *ptprd*, *ptprs*, and *l1cam*), a deeply studied hippocampal synapse involved in associative learning and spatial memory in mice (Nakazawa et al. 2004; Soltesz and Losonczy 2018). In mice, estrogen increases the magnitude of long-term potentiation at this synapse (C. C. Smith, Vedder, and McMahon 2009). It is interesting to speculate that estrogen may regulate plasticity at a conserved hippocampal circuit motif during castle-building behavior. Second, building-associated increases in pNG expression were strongest in populations defined by neuromodulatory receptor genes, and were stronger in populations defined by ERs (*esr1*, *esr2*, *esrra*, *esrrb*, *esrrg*) compared to other receptor families, consistent with previous reports of estrogen-mediated neural plasticity in the mammalian forebrain (Barha and Galea 2010; Brinton 2009; Srivastava and Penzes 2011). Third, building was associated with strong increases in aromatase expression in RGCs, an effect that was driven most strongly by RGC_3. These data thus identify a molecular and cellular pathway that may coordinate building-associated effects of estrogen on brain gene expression, neural circuit structure and function, and male social behavior, consistent with previous work demonstrating estrogenic regulation of male social behaviors in diverse lineages (M. V. Wu et al. 2009; Huffman, O’Connell, and Hofmann 2013; Sonoko Ogawa et al. 2020; Ervin et al. 2015).

An evolutionarily divergent gene module links stem-like glia to neuronal rebalancing and behavior

Appreciation of the importance of glial cells is surging in behavioral neuroscience, with a growing body of work indicating their fundamental roles in synaptic communication, plasticity, learning, memory, behavior, and psychiatric disease (Santello, Toni, and Volterra 2019; Kastanenka et al. 2020; Nagai et al. 2021; Yu et al. 2018). In addition to building-associated aromatase expression in RGCs, we observed building-associated changes in RGC subpopulation-specific gene expression, relative proportions, and signatures of quiescence. Remarkably, comparative genomic analyses across a large number of castle-building, pit-digging, and outgroup rock-dwelling species further converged on the importance of RGCs in castle-building behavior, raising the possibility that transcriptional specializations in glia have served as a primary substrate in castle-building evolution. A module of 12 CDGs showed especially strong co-expression and mapped to RGC subpopulations bearing transcriptional signatures of RGC “stem-like” quiescence. Expression of this module was tightly linked with expression of genetic markers of RGC quiescence and *npas3*, a TF that regulates proliferation of human glia (Moreira et al. 2011). Remarkably, building was associated with downregulation of

both the CDG module (particularly *wdr73*), markers of quiescence, and *npas3* in RGC_2. Furthermore, lasso, elastic net, and ridge regression combined with mediation analysis identified building-associated decreases in expression of *wdr73* as the strongest predictor and the only significant mediator of building-associated neuronal rebalancing. Interestingly, one study in human epithelial cells found that suppressed *wdr73* expression was most strongly associated with increased expression of *ccnd1* (Tilley et al. 2021), an established marker of proliferation in RGCs/neural stem cells in vertebrates (Lukaszewicz and Anderson 2011; Zhang et al. 2021). Spatial profiling further showed that both neuronal populations localized to Dl, the putative fish homologue of the hippocampus. Together, these data support a model whereby castle-building evolved in part by modifying gene regulatory networks that control RGC quiescence and cell type-specific neurogenesis within the putative fish hippocampus.

The CDG module resides in a 19 Mbp genomic region that exhibits signals of divergence mirroring those reported for chromosomal inversions in other species systems (Lamichhaney et al. 2016; da Silva et al. 2019; Tuttle et al. 2016; Corbett-Detig and Hartl 2012; Roesti et al. 2015; Maney et al. 2020; Berg et al. 2017). It is thought that inversions can facilitate rapid evolution by protecting large-scale and adaptive cis-regulatory landscapes and multi-allele haplotypes ("supergenes") from recombination (Schaal, Haller, and Lotterhos 2022; Hoffmann and Rieseberg 2008; Kirkpatrick and Barton 2006; Villoutreix et al. 2021). Evidence for the importance of inversions in phenotypic evolution has been shown in diverse lineages spanning flowers and humans (Huang and Rieseberg 2020; Stefansson et al. 2005). Two recent studies in the ruff and white-throated sparrows further support that inversions may shape social behavioral evolution in diverse lineages (Merritt et al. 2020; Purcell et al. 2014; Küpper et al. 2016). In our data, four genes in the CDG module, including *wdr73*, are immediately proximate to one end of the 19 Mbp region exhibiting strong behavior-associated divergence. It is therefore intriguing to speculate that these genes reside near an inversion "break point" region with a divergent cis-regulatory architecture in castle-building lineages. Future work is needed to determine if an inversion has shaped cis-regulatory expression of these genes, RGC function, and the evolution of castle-building behavior in Lake Malawi cichlid fishes.

LIMITATIONS OF THE STUDY

The molecular readout in this study was nuclear RNA which may not reflect protein function, for example due to post-transcriptional regulation. Because nuclear RNA can only be captured at a single time point within each individual, temporal analysis of decision-making making behaviors during building was limited. This study only profiled the telencephalon, and other brain regions may play critical roles in castle-building. Lastly, firing properties and circuit connections among populations is not determined by snRNA-seq, and thus future experiments are required to determine the behavioral roles of specific neural circuits.

ACKNOWLEDGEMENTS

We thank our collaborators Ashley Parker and Drs. Swantje Graetsch, Manuel Stemmer, and Herwig Baier for valuable feedback during the early stages of the project; Dr. Nicholas Johnson for suggestions regarding statistical analysis of IEG co-expression; Cristina Baker for her critical role in initial development of spatial transcriptomics wetlab pipelines; and the Georgia Tech Petit Institute Genome Analysis and Molecular Evolution Cores for their integral roles in sample processing and sequencing, respectively. This work was supported in part by NIH R01GM101095 to J.T.S., NIH F32GM128346 to Z.V.J., NIH R35 GM139594 to P.T.M., NSF Graduate Research Fellowship DGE-2039655 to T.J.L., and Human Frontiers Science Program RGP0052/2019 to J.T.S.

CONTRIBUTIONS

General: Z.V.J. initially conceived of the experiment and Z.V.J., J.T.S, and B.E.H. developed and designed it. Z.V.J. and B.E.H. performed all wetlab work (see details below under "Wetlab"). T.J.L. pre-processed behavioral and depth data, including in part spatial and temporal registration of both data streams and temporal anchoring to experimental endpoints. G.W.G. pre-processed snRNA-seq, DNA-seq, and spatial transcriptomics data. Z.V.J. and G.W.G. performed downstream data analysis (see details below under "Drylab"). B.E.H. matched snRNA-seq data to published neuroanatomical expression profiles (see details below under "Drylab"). Z.V.J. took the lead on writing the manuscript with critical feedback from G.W.G., J.T.S., B.E.H., and P.T.M. Z.V.J took the lead on designing and creating figures with contributions from

B.E.H., G.W.G., and T.J.L., and with critical feedback from G.W.G., J.T.S., B.E.H., P.T.M., and T.J.L. J.T.S. mentored and funded Z.V.J., B.E.H., and G.W.G., and P.T.M mentored and funded T.J.L on the project. J.T.S. funded snRNA-seq, DNA-seq, and spatial transcriptomics experiments.

Wetlab: Z.V.J. and B.E.H. developed and optimized a single nucleus isolation protocol for cichlid telencephala. Z.V.J. and B.E.H. performed all behavioral assays, surgeries, and downstream nuclei isolations for snRNA-seq. Z.V.J. performed DNA isolations for matching nuclei to subjects. B.E.H. performed all behavior assays for spatial transcriptomics. Z.V.J. and B.E.H. performed surgeries for spatial transcriptomics. B.E.H. performed all downstream wetlab work for spatial transcriptomics. The Petit Institute Genome Analysis Core at GT performed library preparation for snRNA-seq, DNA-seq, and spatial transcriptomics. The Petit Institute Molecular Evolution Core at GT performed sequencing for snRNA-seq, DNA-seq, and spatial transcriptomics.

Drylab: Z.V.J. performed clustering and cluster marker analysis. B.E.H. systematically surveyed the literature to determine conserved neuroanatomical expression patterns of ligand, receptor, nTF, and other cell type-specific marker genes in the teleost telencephalon. B.E.H., G.W.G, and Z.V.J. collaboratively identified markers of RGC quiescence, cycling, and neuronal differentiation. Z.V.J. and G.W.G. collaboratively developed many analytical approaches. Z.V.J. conducted behavioral, IEG co-expression, IEG, DEG, pNG, cell proportion, and gene set enrichment (for biological categories) analyses. G.W.G. matched nuclei to subjects and conducted comparative genomics, gene orthologue calling, ERE detection, gene module detection, and cluster enrichment (for gene lists) analyses. G.W.G. performed spatial integration of clusters and B.E.H. matched spatial transcriptomic profiles to brain regions.

STAR METHODS

EXPERIMENTAL MODEL AND SUBJECT DETAILS

All cichlids (species *Mchenga conophoros*) used in this study were fertilized and raised into adulthood (>180 days) in the Engineered Biosystems Building cichlid aquaculture facilities at the Georgia Institute of Technology in Atlanta, GA in accordance with the Institutional Animal Care and Use Committee guidelines (IACUC protocol number A100029). This colony was originally derived from wild-caught populations collected in Lake Malawi. All experimental animals were collected as fry at approximately 14 days post-fertilization from mouthbrooding females and were raised with broodmates on a ZebTec Active Blue Stand Alone system. At approximately 60 days post-fertilization, animals were transferred to 190-L (92 cm long x 46 cm wide x 42 cm tall) glass aquaria and were housed in social communities (20-30 mixed-sex individuals) into adulthood. Environmental conditions of aquaria were similar to those of the Lake Malawi environment; subjects were maintained on a 12-h:12-h light:dark cycle with full lights on between 8am-6pm Eastern Standard Time (EST) and dim lights on for 60 minutes between light-dark transition (7am-8am and 6pm-7pm EST) in pH=8.2, 26.7°C water and fed twice daily (Spirulina Flake; Pentair Aquatic Ecosystems, Apopka, FL, U.S.A.). All tanks were maintained on a central recirculating system. Reproductive adult subject males (age 6-14 months post-fertilization, n=38) were visually identified from home tanks based on nuptial coloration and expression of classic courtship behaviors (i.e. chasing/leading, quivering). Reproductive adult stimulus females were visually identified from home tanks based on distension of the abdomen (caused by ovary growth) and/or buccal cavity (caused by mouthbrooding).

METHOD DETAILS

Behavior tanks

Behavior tanks were equipped with LED strip lighting synced with external room lighting to minimize large shadows and maximize consistency in video data used for action recognition (10-h:14-h light:dark cycle). Sand (Sahara Sand, 00254, Carib Sea Inc.; ACS00222) was contained within a 38.1 cm long x 45.6 cm wide section of each tank and separated from the rest of the aquarium by a custom 45.6 cm wide x 17.8 cm tall x 0.6 cm thick transparent acrylic barrier secured with plastic coated magnets (1.25 cm wide x 2.5 cm tall x 0.6 cm thick; BX084PC-BLK, K&J Magnetics, Inc.). This design ensured that all fish could freely enter and leave the enclosed sand tray region throughout the trial. At the start of the trial, the smoothed sand surface lay approximately 29.5 cm directly below a custom-designed transparent acrylic tank cover (38.1 cm long x 38.1 cm wide x 3.8 cm tall) that directly contacted the water surface to eliminate rippling for top-down depth sensing and video recordings.

Behavior assays

Subject males were introduced to behavioral tanks containing sand and four reproductive adult age- and size-matched stimulus females of the same species. Broods were collected from all mouthbrooding females prior to introduction of subject males to behavior tanks. Prior to behavioral trials, each male was allowed to initiate castle-building to 1) confirm capacity and motivation to build and 2) minimize potential confounding effects of "novelty" on brain gene expression that may be caused by the male's first experience building. After building was confirmed during the initial "pre-trial" period, the sand surface in each behavioral tank was smoothed shortly before lights off, and an automated depth sensing and video recording protocol was initiated as previously described using a Raspberry Pi 3 mini-computer (Raspberry Pi Foundation) (Johnson, Arrojwala, et al. 2020). Briefly, this system uses 1) a Microsoft XBox Kinect Depth sensor to track depth change across the sand surface every five minutes, enabling analysis of the developing bower structure over time, and 2) a Raspberry Pi v2 camera to record 10 hours of high-definition video data daily. The system regularly uploads depth change updates to a Google Documents spreadsheet, enabling real-time, remote monitoring of bower construction activity in each tank. Following each trial, a trained 3D Residual Network was used to predict male building and quivering behaviors from video data as previously described (Long et al. 2020).

Tissue sampling

Actively constructing males (n=19) were identified through remote depth change updates and were collected between 11am-2pm EST (3-5 h after full lights-on and feeding) to control for potential effects of circadian rhythm, feeding, hunger, and anticipation of food on brain gene expression. At the same time, a neighboring male that was not constructing a bower (nor had initiated construction) but could also freely interact with four females and sand, was also collected ("control", n=19). Immediately following collection, subjects were rapidly anesthetized with tricaine for rapid brain extraction, measured for standard length (SL, distance measured from snout to caudal peduncle) and body mass (BM), and rapidly decapitated for brain extraction. Telencephala (including olfactory bulbs) were dissected under a dissection microscope (Zeiss Stemi DV4 Stereo Microscope 8x - 32x, 000000-1018-455), in Hibernate AB Complete nutrient medium (HAB; with 2% B27 and 0.5 mM Glutamax; BrainBits) containing 0.2 U/µl RNase Inhibitor (Sigma). Immediately following dissection telencephala were rapidly frozen on powdered dry ice and stored at -80 °C. Testes were then surgically extracted and weighed to calculate gonadosomatic index (GSI=gonad mass/BM*100) for each subject (subject information available in Table S8).

Nuclei isolation

Nuclei were isolated following a protocol adapted from (Martelotto 2020) and optimized for cichlid telencephala.

snRNA-sequencing

Suspensions of isolated nuclei were loaded onto the 10x Genomics Chromium Controller (10x Genomics) at concentrations ranging from 400-500 nuclei/ul with a target range of 3,000–4,000 nuclei per sample. Downstream cDNA synthesis and library preparation using Single Cell 3' GEM, Library and Gel Bead Kit v3.1 and Chromium i7 Multiplex Kit were performed according to manufacturer instructions (Chromium Single Cell 3' Reagent Kits User Guide v3.1 Chemistry, 10X Genomics). Sample quality was assessed using high sensitivity DNA analysis on the Bioanalyzer 2100 system (Agilent) and libraries were quantified using a Qubit 2.0 (Invitrogen). Barcoded cDNA libraries were pooled and sequenced on the NovaSeq 6000 platform (Illumina) on a single flow cell using the 300-cycle S4 Reagent kit (2x150 bp paired-end reads; Illumina).

DNA sequencing

Genomic DNA was isolated from diencephalic tissue sampled from each test subject using a DNeasy Blood and Tissue Kit pipeline with a 60 min lysis time and without RNase A. The 260/280 nm absorbance ratio ranged from 1.91-2.10 across subjects. Libraries were prepared following a NEBNext Ultra II FS DNA Library Prep Kit for Illumina protocol. Libraries were sequenced on two NovaSeq 6000 lanes using 300-cycle SP Reagent Kits (2x150 bp paired-end reads; Illumina).

Spatial transcriptomics

Telencephala were microdissected from two size-matched build-control pairs of MC males (n=4 males total), embedded in cryomolds, flash frozen on dry ice, and stored at -80°C until further processing. Tissue was cryo-sectioned coronally at 10-µm thickness at -20°C (Cryostar NX70) and mounted onto pre-chilled Visium Spatial Gene Expression slides (10X Genomics). RNA quality (RIN > 7) was confirmed using an RNA 6000 Nano Kit (Agilent). Spatial gene expression slides were processed following manufacturer instructions (Visium Spatial Gene Expression Reagent Kits User Guide, 10X Genomics). Barcoded cDNA libraries were sequenced on the NovaSeq 6000 platform (Illumina).

QUANTIFICATION AND STATISTICAL ANALYSIS

Behavioral Analysis

For all trials, 3D ResNet-predicted behaviors and structural change across the sand surface was analyzed over the 90 minutes preceding collection following the same general approach described previously (Johnson, Arrojwala, et al. 2020). Briefly, a smoothing algorithm was applied to remove depth change

attributable to technical noise, and small regions of missing data were recovered by spatial interpolation. Bowers were defined as any region within which one-thousand or more contiguous pixels (equivalent to ~10 cm^2) changed in elevation by more than 0.2 cm in the same direction (~2 cm^3 volume change total) based on previous analysis of depth change caused by non-building home tank activity (Johnson, Arrojwala, et al. 2020). Depth change values were adjusted based on the cubed standard length of each subject male, to create a standardized measure of building activity (larger males have larger mouths and can scoop and spit a larger volume of sand). Action recognition was used to track the number, location, and timepoints of predicted bower construction behaviors (scoops, spits, and multiple events) and quivering behaviors over the same 90 min period. The number of quivering events was log-normalized due to a single outlier (building) male with 257 predicted quivering events (~5.9 standard deviations above the mean). Feeding behaviors were not analyzed because they can be performed by both males and females and we are not able to reliably attribute individual feeding events to the subject male.

For simplicity, we generated a single "Bower Activity Index" (BAI) metric to reflect overall building activity by first calculating the regression line between depth change and building events for each trial (n=38, R^2=0.76). We then projected each male's depth change and bower behavior values onto that line, with the lowest value (0 predicted building events, 0 above threshold depth change) being set to 0. BAI was calculated as the Euclidean distance along the regression line from the lowest value. BAI was used as a continuous measure of castle-building behavior throughout this study.

Differences in building, quivering, and GSI between groups were analyzed using a paired t-test in which behave and control subjects collected at the same time were treated as pairs.

snRNA-seq pre-processing and quality control

FASTQ files were processed with Cell Ranger version 3.1.0 (10X Genomics). Reads were aligned to the *Maylandia zebra* Lake Malawi cichlid genome assembly (Conte et al. 2019) using a splice-aware alignment algorithm (STAR) within Cell Ranger, and gene annotations were obtained from the same assembly (NCBI RefSeq assembly accession: GCF_000238955.4, M_zebra_UMD2a). Because nuclear RNA contains intronic sequence, there were included in the cellranger count step. Cell Ranger filtered out UMIs that were homopolymers, contained N, or contained any base with a quality score less than 10. Following these steps, Cell Ranger generated ten filtered feature-barcode matrices (one per pool) containing expression data for a total of 32,471 features (corresponding to annotated genes) and a total of 33,895 barcodes (corresponding to droplets and putative nuclei) that were used passed through additional quality control steps in the "Seurat" package in R. Examination of total transcripts, total genes, and proportion of mitochondrial transcripts were similar across all ten pools, and therefore the same criteria were used to remove potentially dead or dying nuclei from all pools. Barcodes associated with fewer than 300 total genes, fewer than 500 total transcripts, or greater than 5% (of total transcripts) mitochondrial genes were excluded from downstream analysis on this basis. This step filtered out a total of 20 (0.059%) barcodes. To reduce risk of doublets or multiplets, barcodes associated with more than 3,000 total genes or 8,000 total transcripts were also excluded. This step filtered out a total of 201 barcodes (0.59%). In total, 33,674 barcodes (99.34%) passed all quality control filters and were included in downstream analyses.

Dimensionality reduction

In order to perform dimensionality reduction, we first identified 4,000 genes that exhibited the most variable expression patterns across nuclei using the FindVariableFeatures function in Seurat with the mean.var.plot selection method, which aims to identify variable features while controlling for the strong relationship between variability and average expression, and otherwise default parameters. Gene-level data was then scaled using the ScaleData function in Seurat with default parameters. To examine dimensionality, we first performed a linear dimensional reduction using the RunPCA command with the maximum possible number of dimensions ("dim" set to 50). We then used Seurat's JackStraw, ScoreJackStraw, and JackStrawPlot functions to estimate and visualize the significance of the first 50 principal components (PCs), and the Elbow plot function to visualize the variance explained by the first 50 PCs. Because all 50 PCs were highly statistically significant, and no "drop off" was observed in variance explained across PCs, we used all 50 PCs for non-linear dimensional reduction (Uniform Manifold Approximation and Projection, UMAP) using the RunUMAP function

in Seurat. For RunUMAP, “min.dist” was set to 0.5, “n.neighbors” was set to 50, and “metric” was set to “euclidean”.

Clustering

Prior to clustering, nuclei were embedded into a K nearest-neighbor (KNN) graph based on euclidean distance in UMAP space, with edge weights based on local Jaccard similarity, using the FindNeighbors function in Seurat (k.param=50, dims=1:2, prune.SNN=0). Clustering was then performed using Seurat’s FindClusters function using the Louvain algorithm with multilevel refinement (algorithm=2). This final step was performed twice using two different resolution parameters to generate both coarse- and fine-grained structural descriptions of the underlying data, facilitating investigation of both major cell types as well as smaller subpopulations. For more coarse-grained clustering (resolution=0.01) identified 15 1° clusters and fine-grained clustering (resolution=1.3) identified 53 2° clusters.

Cluster marker gene analysis

The biological identities of specific clusters were investigated using a multi-pronged approach that incorporated unbiased analysis of cluster-specific marker genes as well as supervised examination of previously established marker genes. Cluster-specific “marker” genes were identified using the FindAllMarkers function in Seurat. Briefly, this function compares gene expression within each cluster to gene expression across all other clusters and calculates Bonferroni-adjusted p-values using a Wilcoxon rank sum test. Functional enrichment analysis of GO categories among cluster-specific marker genes was investigated by first converting cichlid gene names to their human orthologs and then performing functional enrichment analysis using ToppGene Suite with default parameters . Enrichment results that survived FDR-adjustment ($q<0.05$) were considered statistically significant. Established cell type-specific and neuroanatomical marker genes were identified from the literature (Table S2) and were intersected with the output from FindAllMarkers to generate further insight into the biological identity of clusters.

Assignment of nuclei to test subjects

To match individual nuclei to individual test subjects, we used Demuxlet to match variants identified in snRNA-seq reads to variants identified from genomic sequencing of each subject (Kang et al. 2018). First, genomic DNA from every test subject was collected and sequenced. In total, 276.7 Gbp of sequenced reads were assigned quality scores≥30 (91.4% of all reads). The corresponding FASTQ files were filtered and aligned to the M. zebra Lake Malawi cichlid genome umd2a assembly (NCBI RefSeq assembly accession: GCF_000238955.4, M_zebra_UMD2a). The resulting bam file was sorted, duplicates removed, read groups added, and indexed using Picard tools. Variants were then called using GATK v4.1.8.1 HaplotypeCaller using the M. zebra umd2a reference genome. Based on pool, individual vcf files were merged, resulting in 10 files (one for each pool). These files were then filtered to retain only variants that varied among individuals in a pool. For each pool, only SNPs for which 1) at least one individual from the pool had a different genotype from the other individuals, and 2) no individuals had missing data, were used as input to Demuxlet. The number of SNPs used ranged from 112,385 to 357,177 with a mean of 241,780±22,369 per pool.

Next, variants were called from snRNA-seq reads following a similar pipeline. Reads from Cell Ranger's output bam file were filtered for those that passed the quality control metrics described above using samtools v1.11. The resulting bam file was sorted, duplicates removed, read groups added, and indexed using Picard tools. Variants were then called using GATK HaplotypeCaller using the M. zebra umd2a reference genome and without the MappingQualityAvailableReadFilter to retain reads that were confidently mapped by Cell Ranger (MAPQ score of 255). The SNPs from the snRNA-seq reads and the genomic DNA were input to Demuxlet, which computed a likelihood estimation that each nucleus belongs to each individual in the pool. Nuclei were assigned to the individual test subject with the greatest probability estimated by Demuxlet.

Identification of IEG-like genes

Three canonical IEGs (*c-fos, egr1, npas4*) were used to identify additional genes exhibited IEG-like expression across clusters. For each of these three IEGs, nuclei were split into IEG-positive versus IEG-negative nuclei within each of the 53 clusters. Within each cluster, differential gene expression was analyzed between IEG-positive versus IEG-negative nuclei using the FindMarkers function in Seurat, with "logfc.threshold" set to 0, and "min.pct" set to 1/57 (57 was selected as this was the number of nuclei in the smallest cluster). Within each cluster, any genes that did not meet this criterion were excluded and were assigned a p-value of 1. Because FindMarkers requires at least three nuclei to be present in both comparison groups, clusters that contained less than three IEG-positive nuclei were excluded. Genes that were detected in the majority of clusters, and that were significantly ($p<0.05$) upregulated in IEG-positive nuclei in the majority of those clusters were considered to be significantly co-expressed with each individual IEG. Genes that were significantly co-expressed with all three IEGs were used as IEG-like markers for downstream analyses of IEG-like expression.

Differential IEG expression

Building-, quivering-, and gonadal-associated IEG expression was analyzed in 1° and 2° clusters, gene-defined populations within 1° and 2° clusters, and gene-defined populations regardless of cluster. To do this, we calculated an IEG score for each nucleus, equal to the number of unique IEG-like genes (n=25) expressed. Building-, quivering-, and gonadal-associated differences in IEG score were analyzed using a beta-binomial model in which the number of IEG-like genes observed as well as the number of the IEG-like genes not observed were tracked as indicators of recent neuronal excitation. This analysis was performed using the 'BBmm' package in R (m=25). Because castle-building, quivering, and GSI were correlated with one another, we analyzed expression using a sequence of beta-binomial mixed-effects models in which different pairwise combinations of predictor variables (building, quivering, and GSI) competed to explain variance in IEG score. These models also included nested random terms to account for variance explained by individual variation, pair, pool, and RNA isolation/cDNA library generation batch. Within this framework, we ran the following seven models, which allowed building (analyzed as either a binary or a continuous variable), quivering, and GSI to compete in all possible combinations to explain variance in IEG score:

1. IEG score ~ **BAI** + **log(quivering events)** + *(subject/pool/batch)* + *(subject/pair)*
2. IEG score ~ **BAI** + **GSI** + *(subject/pool/batch)* + *(subject/pair)*
3. IEG score ~ **BAI** + **log(quivering events)** + **GSI** + *(subject/pool/batch)* + *(subject/pair)*
4. IEG score ~ **Condition** + **log(quivering events)** + *(subject/pool/batch)* + *(subject/pair)*
5. IEG score ~ **Condition** + **GSI** + *(subject/pool/batch)* + *(subject/pair)*
6. IEG score ~ **Condition** + **log(quivering events)** + **GSI** + *(subject/pool/batch)* + *(subject/pair)*
7. IEG score ~ **log(quivering events)** + **GSI** + *(subject/pool/batch)* + *(subject/pair)*

We defined significant building-, quivering-, and gonadal-associated IEG effects as those in which 1) the raw p-value for the corresponding fixed effect (for building, BAI and condition; for quivering, log-normalized quivering; for gonadal, GSI) was significant ($p<0.05$) in every model, and 2) the harmonic mean p-value across models was significant after adjusting for multiple comparisons for all genes and populations analyzed ($hmp_{adj}<0.05$). To calculate the harmonic mean p-value, we used the "harmonicmeanp" package in R. Thus, building-associated IEG effects were significant (the raw p-value for the effect of "condition" and "BAI" <0.05) in models 1-6, and if the harmonic mean p-value across models 1-6 was significant after adjusting for multiple comparisons across all cell populations.

Building-, quivering-, and gonadal-associated gene expression

Building-, quivering-, and gonadal-associated gene expression was analyzed within 1° and 2° clusters using a multiple linear mixed-effects regression approach with the "glmmSeq" package in R (https://github.com/KatrionaGoldmann/glmmSeq). Because castle-building, quivering, and GSI were correlated with one another, we analyzed expression using a sequence of linear mixed-effects regression models in which different pairwise combinations of predictor variables (building, quivering, and GSI) competed to explain variance in gene expression. These models also included nested random terms to account for variance explained by individual variation, pair, sample pool, and 10X Chromium run. Thus, sample size was equal to the number of individuals (n=38), with many repeated observations being recorded

from each individual (equal to the number of nuclei sampled from that individual as assigned to the cluster being analyzed). Building was analyzed both as a continuous variable (BAI) and as a binary categorical variable (behave versus control).

We defined bDEGS, qDEGs, and gDEGs as genes (within clusters) in which expression was significantly (raw p<0.05) associated with the corresponding fixed effect (for bDEGs, BAI and condition; for qDEGs, log-normalized quivering; for gDEGs, GSI) in every model, and additionally in which the harmonic mean p-value across models was significant after adjusting for multiple comparisons for all genes and all clusters (5% false discovery rate). For each model, dispersion was estimated for each gene using the "DESeq2" package in R, using parameters recommended for single cell datasets (fitType = "glmGamPoi", minmu = 1e-06). Size factors for each gene were calculated using the "scran" package in R, using default parameters, except that max.cluster.size was set to the number of nuclei assigned to the cluster being analyzed. Genes that were not observed in 19/19 pairs were excluded from analysis.

Estrogen response element detection

Estrogen receptors are hormone-dependent transcription factors capable of regulating target gene expression by binding to specific DNA sequences called estrogen response elements (EREs). EREs can be easily identified by their prototypic motif of AGGTCA separated by a 3-base spacer (Ikeda, Horie-Inoue, and Inoue 2015). Genes with an ERE motif less than 25 kilobases away were found and the location of the ERE relative to the gene was recorded as either intragenic (ERE within the start site to the 3' polyA tail), promoter (ERE <= 5 kb upstream of the gene), or distal (all other locations less than 25 kb away from the gene). To identify the location of the ERE to the closest gene, bedtools v2.29.1 was used using the closest command.

Building-, quivering-, and gonadal-associated pNG expression

Building-, quivering-, and gonadal-associated pNG expression was analyzed in 1° and 2° clusters, gene-defined populations within 1° and 2° clusters, and gene-defined populations regardless of cluster using the same general approach described for IEG expression, except that building-associated effects were defined as those that were significantly associated with condition in all models. Because we did not expect neurogenesis or associated cellular processes to proceed over 90-minute timescales, we did not additionally require effects to be significantly associated with BAI in all models.

Building-associated changes in cell proportions

Behavior-associated differences in cell type-specific proportions were analyzed for 1° and 2° clusters with a binomial mixed-effects regression model using the glmer function within the "lmer" package in R. The model included condition, GSI, and quivering as fixed effects, and included a random term for individual variation. 1° cluster proportions were calculated as the proportion of all nuclei assigned to each 1° cluster, and 2° cluster proportions were calculated as the proportion of 1° "parent" cluster nuclei assigned to each 2° "daughter" cluster. Thus, each nucleus was treated as an observation with a binary outcome (either an instance of the target cluster or not) from an individual that could be explained by condition, quivering, or GSI. p-values were estimated using the 'lmerTest' package in R, and qvalues were calculated using the 'qvalue' package in R. Building-associated effects were defined as those that were significant after accounting for multiple comparisons across all clusters with a false discovery rate of 5% (q<0.05).

Cluster-specific enrichment of gene sets

To test if genes associated with the evolution of bower construction behavior (identified through comparative genomics) were enriched in specific cell populations, we first calculated a "gene set score" for each nucleus, equal to the total number of unique behavioral evolution genes expressed. Because the gene set score could be impacted by the total volume of sequence data sampled from each nucleus, we divided the gene set score by the total number of unique genes expressed in each nucleus. To quantify enrichment, a Z-test was then used to compare "normalized" gene set scores for all nuclei within a cluster compared to all other nuclei. The distribution of the normalized values was assumed to be normal according to the central limit theorem and population standard deviation was approximated using sample standard deviation.

Secondly, the effect size, as measured by Cohen's d, of the results were compared to those of random gene lists. To prevent differences in overall amount of expression between random genes and genes of interest from skewing results, random genes lists were chosen that had approximately equal number of UMIs expressed as a whole to the genes of interest. This was achieved by first ranking all the genes from the highest number of UMIs expressed to the lowest. Next, for each gene of interest, a pool of 100 random genes were chosen that were ranked most closely to the gene of interest and were not a gene of interest themselves. Then, 10,000 random gene lists were created by choosing one gene at random from each pool. The enrichment test described above was then performed on the random gene lists. Finally, clusters that were significantly enriched for the genes of interest according to the process above and had significantly greater effect sizes than the 10,000 random gene lists were considered to be significant.

RGC subclustering

RGC subclusters were determined using the same general procedure used for clustering 1° and 2° clusters but restricted to only those nuclei assigned to 1.1_RGC and 1.2_RGC.

Analysis of castle-associated genomic divergence

In order to identify potential behavior-associated genomic variants, comparative genomic analyses were performed using genomic sequence data collected from 27 Lake Malawi cichlid species (Patil et al. 2021). Fixation indices (F_{ST}) were calculated for polymorphic variants in two separate analyses using vcftools v0.1.17. The first analysis compared pit-digging (N=11) versus castle-building (N=9) species, and the second compared rock-dwelling (N=7) versus castle-building (N=9) species. Variants for which sequence data was missing from 50% or more of species in either group were excluded from analysis. F_{ST} analyses were performed separately using the --weir-fst-pop and --fst-window-size 10000 flag to calculate F_{ST} across 10kb bins in vcftools. Then, bins where F_{ST} was greater than 0.20 in the pit-castle comparison and 0.20 in the rock-castle comparison were kept. These thresholds are both greater than the minimum F_{ST} of FDR-adjusted significant bins. By creating these more strict thresholds we aimed to ensure that the selected bins were extremely divergent between castle-building and non-castle-building species. Additionally, a higher threshold was selected for the rock-castle than the pit-castle comparison because of the greater evolutionary distance and thus greater overall F_{ST}. Finally, genes that were within 25kb of these bins meeting these thresholds were identified using bcftools v1.11 with the closest command and the *M. zebra* genome as reference. Genes within 25kb of highly divergent pit-castle and rock-castle bins are referred to here as "castle-divergent".

CDG co-expression and module analysis

Modules of co-expressed CDGs were analyzed using weighted correlation network analysis (WGCNA) using the “WGCNA” v1.70-3 package in R. CDGs that were not observed in any nucleus were excluded from analysis. The normalized gene expression data for CDGs was used as the input gene expression matrix and the function pickSoftThreshold was used to determine the optimal soft-thresholding power. We determined the optimal soft-thresholding power to be 1 because it was the lowest power for which the scale-free topology fit index reached 0.90. Then an adjacency matrix was created from the input gene expression matrix using the adjacency function with power = 1, type = "signed" and otherwise default parameters. The adjacency matrix was used as the topological overlap matrix (TOM) and the dissimilarity matrix was calculated as 1 – TOM. To detect modules, k-means clustering was performed using all possible values of k and the results were compared to determine the optimal k. First, a distance matrix was constructed from the dissimilarity matrix produced by WGNCA using the dist function in R. Next, the function pam from the R package “cluster” v2.1.0 was used to cluster the distance matrix with diss = T, otherwise default parameters, and k set to the value that produced the highest average silhouette width for all genes. Briefly, silhouette width is a measure of the similarity of the genes within a module to the genes outside the module, and higher values indicate better clustering. We found that k=2, had the greatest average silhouette width. The strength of the module was evaluated using a two-sampled Welch t-test comparing the silhouette width and gene-gene correlations for CDGs within the module versus CDGs outside the module. To analyze the relationship between the CDG module and signatures of RGC quiescence, the correlation coefficient was calculated based the number of genes in the CDG module expressed in each nucleus versus the number of quiescent markers expressed in

each nucleus. We compared the correlation coefficient against a permuted null distribution by randomly shuffling the expression values of each gene in the module 10,000 times.

Spatial transcriptomic pre-processing and quality control

Base Call files were demultiplexed into FASTQ files and processed with Space Ranger v1.3.1 (10X Genomics). Reads were aligned to the *M. zebra* umd2a reference assembly as described above for snRNA-seq (Conte et al. 2019). Following these steps, Space Ranger generated three filtered feature-barcode matrices containing expression data for a total of 32,471 features (corresponding to annotated genes). Spots with 0 UMIs were removed resulting in 6,707 spots used in downstream analysis.

Spatial integration of snRNA-seq clusters

To predict locations of specific snRNA-seq identified clusters in spatial transcriptomics data, an 'anchor'-based integration workflow in Seurat was used. First, both the snRNA-seq and spatial data were normalized using the SCTransform in Seurat. Next, anchors were identified between the reference snRNA-seq and the query spatial data using FindTransferAnchors in Seurat, and a matrix of predictions scores was generated for each cluster in every spot using the TransferData function in Seurat. The maximum prediction score across clusters was not uniform, therefore we normalized the values between 0 and 1 in order to enable meaningful comparisons across cell types.

REFERENCES

Adkins-Regan, Elizabeth. 2013. *Hormones and Animal Social Behavior*. Princeton University Press.

Adolf, Birgit, Prisca Chapouton, Chen Sok Lam, Stefanie Topp, Birgit Tannhäuser, Uwe Strähle, Magdalena Götz, and Laure Bally-Cuif. 2006. "Conserved and Acquired Features of Adult Neurogenesis in the Zebrafish Telencephalon." *Developmental Biology* 295 (1): 278–93.

Almeida, Olinda, Ana S. Félix, Gonçalo A. Oliveira, João S. Lopes, and Rui F. Oliveira. 2019. "Fighting Assessment Triggers Rapid Changes in Activity of the Brain Social Decision-Making Network of Cichlid Fish." *Frontiers in Behavioral Neuroscience* 13 (September): 229.

Almli, Lynn M., and Walter Wilczynski. 2012. "Socially Modulated Cell Proliferation Is Independent of Gonadal Steroid Hormones in the Brain of the Adult Green Treefrog (Hyla Cinerea)." *Brain, Behavior and Evolution* 79 (3): 170–80.

Alward, Beau A., Austin T. Hilliard, Ryan A. York, and Russell D. Fernald. 2019. "Hormonal Regulation of Social Ascent and Temporal Patterns of Behavior in an African Cichlid." *Hormones and Behavior* 107 (January): 83–95.

Amadei, Elizabeth A., Zachary V. Johnson, Yong Jun Kwon, Aaron C. Shpiner, Varun Saravanan, Wittney D. Mays, Steven J. Ryan, et al. 2017. "Dynamic Corticostriatal Activity Biases Social Bonding in Monogamous Female Prairie Voles." *Nature* 546 (7657): 297–301.

Amenyogbe, Eric, Gang Chen, Zhongliang Wang, Xiaoying Lu, Mingde Lin, and Ai Ying Lin. 2020. "A Review on Sex Steroid Hormone Estrogen Receptors in Mammals and Fish." *International Journal of Endocrinology* 2020 (February): 5386193.

Anderson, David J. 2016. "Circuit Modules Linking Internal States and Social Behaviour in Flies and Mice." *Nature Reviews. Neuroscience* 17 (11): 692–704.

Asrican, Brent, Josh Wooten, Ya-Dong Li, Luis Quintanilla, Feiran Zhang, Connor Wander, Hechen Bao, et al. 2020. "Neuropeptides Modulate Local Astrocytes to Regulate Adult Hippocampal Neural Stem Cells." *Neuron* 108 (2): 349-366.e6.

Bachevalier, Jocelyne, and Katherine A. Loveland. 2006. "The Orbitofrontal-Amygdala Circuit and Self-Regulation of Social-Emotional Behavior in Autism." *Neuroscience and Biobehavioral Reviews* 30 (1): 97–117.

Badurek, Sylvia, Marilena Griguoli, Aman Asif-Malik, Barbara Zonta, Fei Guo, Silvia Middei, Laura Lagostena, et al. 2020. "Immature Dentate Granule Cells Require Ntrk2/TrkB for the Formation of Functional Hippocampal Circuitry." *IScience* 23 (5): 101078.

Balthazart, Jacques, and Gregory F. Ball. 2016. "Endocrine and Social Regulation of Adult Neurogenesis in Songbirds." *Frontiers in Neuroendocrinology* 41 (April): 3–22.

Baran, Nicole M., and J. Todd Streelman. 2020. "Ecotype Differences in Aggression, Neural Activity and Behaviorally Relevant Gene Expression in Cichlid Fish." *Genes, Brain, and Behavior* 19 (6): e12657.

Barha, Cindy K., and Liisa A. M. Galea. 2010. "Influence of Different Estrogens on Neuroplasticity and Cognition in the Hippocampus." *Biochimica et Biophysica Acta* 1800 (10): 1056–67.

Bedos, M., W. Portillo, and R. G. Paredes. 2018. "Neurogenesis and Sexual Behavior." *Frontiers in Neuroendocrinology* 51 (October): 68–79.

Bendesky, Andres, Young-Mi Kwon, Jean-Marc Lassance, Caitlin L. Lewarch, Shenqin Yao, Brant K. Peterson, Meng Xiao He, Catherine Dulac, and Hopi E. Hoekstra. 2017. "The Genetic Basis of Parental Care Evolution in Monogamous Mice." *Nature* 544 (7651): 434–39.

Berg, P. R., B. Star, C. Pampoulie, I. R. Bradbury, P. Bentzen, J. A. Hutchings, S. Jentoft, and K. S. Jakobsen. 2017. “Trans-Oceanic Genomic Divergence of Atlantic Cod Ecotypes Is Associated with Large Inversions.” *Heredity* 119 (6): 418–28.

Boender, Arjen J., and Larry J. Young. 2020. “Oxytocin, Vasopressin and Social Behavior in the Age of Genome Editing: A Comparative Perspective.” *Hormones and Behavior* 124 (104780): 104780.

Brenowitz, Eliot A., and Harold H. Zakon. 2015. “Emerging from the Bottleneck: Benefits of the Comparative Approach to Modern Neuroscience.” *Trends in Neurosciences* 38 (5): 273–78.

Brinton, Roberta Diaz. 2009. “Estrogen-Induced Plasticity from Cells to Circuits: Predictions for Cognitive Function.” *Trends in Pharmacological Sciences* 30 (4): 212–22.

Chen, Xi, Xiao Li, Yin Ting Wong, Xuejiao Zheng, Haitao Wang, Yujie Peng, Hemin Feng, et al. 2019. “Cholecystokinin Release Triggered by NMDA Receptors Produces LTP and Sound-Sound Associative Memory.” *Proceedings of the National Academy of Sciences of the United States of America* 116 (13): 6397–6406.

Conte, Matthew A., Rajesh Joshi, Emily C. Moore, Sri Pratima Nandamuri, William J. Gammerdinger, Reade B. Roberts, Karen L. Carleton, Sigbjørn Lien, and Thomas D. Kocher. 2019. “Chromosome-Scale Assemblies Reveal the Structural Evolution of African Cichlid Genomes.” *GigaScience* 8 (4). https://doi.org/10.1093/gigascience/giz030.

Corbett-Detig, Russell B., and Daniel L. Hartl. 2012. “Population Genomics of Inversion Polymorphisms in Drosophila Melanogaster.” *PLoS Genetics* 8 (12): e1003056.

Dias, Caroline M., and Christopher A. Walsh. 2020. “Recent Advances in Understanding the Genetic Architecture of Autism.” *Annual Review of Genomics and Human Genetics* 21 (1): 289–304.

Diotel, Nicolas, Colette Vaillant, Cyril Gabbero, Svetlana Mironov, Alexis Fostier, Marie-Madeleine Gueguen, Isabelle Anglade, Olivier Kah, and Elisabeth Pellegrini. 2013. “Effects of Estradiol in Adult Neurogenesis and Brain Repair in Zebrafish.” *Hormones and Behavior* 63 (2): 193–207.

Duarte-Guterman, Paula, Shunya Yagi, Carmen Chow, and Liisa A. M. Galea. 2015. “Hippocampal Learning, Memory, and Neurogenesis: Effects of Sex and Estrogens across the Lifespan in Adults.” *Hormones and Behavior* 74 (August): 37–52.

Dulac, Catherine, Lauren A. O’Connell, and Zheng Wu. 2014. “Neural Control of Maternal and Paternal Behaviors.” *Science (New York, N.Y.)* 345 (6198): 765–70.

Dunlap, Kent D., Michael Chung, and James F. Castellano. 2013. “Influence of Long-Term Social Interaction on Chirping Behavior, Steroid Levels and Neurogenesis in Weakly Electric Fish.” *The Journal of Experimental Biology* 216 (Pt 13): 2434–41.

Elliott, S. Benjamin, Erik Harvey-Girard, Ana C. C. Giassi, and Leonard Maler. 2017. “Hippocampal-like Circuitry in the Pallium of an Electric Fish: Possible Substrates for Recursive Pattern Separation and Completion.” *The Journal of Comparative Neurology* 525 (1): 8–46.

Ervin, Kelsy S. J., Jennifer M. Lymer, Richard Matta, Amy E. Clipperton-Allen, Martin Kavaliers, and Elena Choleris. 2015. “Estrogen Involvement in Social Behavior in Rodents: Rapid and Long-Term Actions.” *Hormones and Behavior* 74 (August): 53–76.

Furth, W. R. van, G. Wolterink, and J. M. van Ree. 1995. “Regulation of Masculine Sexual Behavior: Involvement of Brain Opioids and Dopamine.” *Brain Research. Brain Research Reviews* 21 (2): 162–84.

Gallant, Jason R., and Lauren A. O’Connell. 2020. “Studying Convergent Evolution to Relate Genotype to Behavioral Phenotype.” *The Journal of Experimental Biology* 223 (Pt Suppl 1): jeb213447.

Gangopadhyay, Prabaha, Megha Chawla, Olga Dal Monte, and Steve W. C. Chang. 2021. "Prefrontal-Amygdala Circuits in Social Decision-Making." *Nature Neuroscience* 24 (1): 5–18.

Ganz, Julia, and Michael Brand. 2016. "Adult Neurogenesis in Fish." *Cold Spring Harbor Perspectives in Biology* 8 (7). https://doi.org/10.1101/cshperspect.a019018.

Goodson, James L. 2005. "The Vertebrate Social Behavior Network: Evolutionary Themes and Variations." *Hormones and Behavior* 48 (1): 11–22.

Götz, Magdalena, and Wieland B. Huttner. 2005. "The Cell Biology of Neurogenesis." *Nature Reviews. Molecular Cell Biology* 6 (10): 777–88.

Gutzeit, Vanessa A., Kylia Ahuna, Tabia L. Santos, Ashley M. Cunningham, Meghin Sadsad Rooney, Andrea Muñoz Zamora, Christine A. Denny, and Zoe R. Donaldson. 2020. "Optogenetic Reactivation of Prefrontal Social Neural Ensembles Mimics Social Buffering of Fear." *Neuropsychopharmacology: Official Publication of the American College of Neuropsychopharmacology* 45 (6): 1068–77.

Guzowski, John F., Jerilyn A. Timlin, Badri Roysam, Bruce L. McNaughton, Paul F. Worley, and Carol A. Barnes. 2005. "Mapping Behaviorally Relevant Neural Circuits with Immediate-Early Gene Expression." *Current Opinion in Neurobiology* 15 (5): 599–606.

Hasenpusch-Theil, Kerstin, Stephen West, Alexandra Kelman, Zrinko Kozic, Sophie Horrocks, Andrew P. McMahon, David J. Price, John O. Mason, and Thomas Theil. 2018. "Gli3 Controls the Onset of Cortical Neurogenesis by Regulating the Radial Glial Cell Cycle through Cdk6 Expression." *Development (Cambridge, England)* 145 (17): dev163147.

Heinrichs, Markus, and Jens Gaab. 2007. "Neuroendocrine Mechanisms of Stress and Social Interaction: Implications for Mental Disorders." *Current Opinion in Psychiatry* 20 (2): 158–62.

Hofer, Sonja B., Thomas D. Mrsic-Flogel, Tobias Bonhoeffer, and Mark Hübener. 2009. "Experience Leaves a Lasting Structural Trace in Cortical Circuits." *Nature* 457 (7227): 313–17.

Hoffmann, Ary A., and Loren H. Rieseberg. 2008. "Revisiting the Impact of Inversions in Evolution: From Population Genetic Markers to Drivers of Adaptive Shifts and Speciation?" *Annual Review of Ecology, Evolution, and Systematics* 39 (1): 21–42.

Hoke, Kim L., Michael J. Ryan, and Walter Wilczynski. 2005. "Social Cues Shift Functional Connectivity in the Hypothalamus." *Proceedings of the National Academy of Sciences of the United States of America* 102 (30): 10712–17.

Howe, Kerstin, Matthew D. Clark, Carlos F. Torroja, James Torrance, Camille Berthelot, Matthieu Muffato, John E. Collins, et al. 2013. "The Zebrafish Reference Genome Sequence and Its Relationship to the Human Genome." *Nature* 496 (7446): 498–503.

Huang, Kaichi, and Loren H. Rieseberg. 2020. "Frequency, Origins, and Evolutionary Role of Chromosomal Inversions in Plants." *Frontiers in Plant Science* 11 (March): 296.

Huffman, Lin S., Lauren A. O'Connell, and Hans A. Hofmann. 2013. "Aromatase Regulates Aggression in the African Cichlid Fish Astatotilapia Burtoni." *Physiology & Behavior* 112–113 (March): 77–83.

Hung, Lin W., Sophie Neuner, Jai S. Polepalli, Kevin T. Beier, Matthew Wright, Jessica J. Walsh, Eastman M. Lewis, et al. 2017. "Gating of Social Reward by Oxytocin in the Ventral Tegmental Area." *Science (New York, N.Y.)* 357 (6358): 1406–11.

Ikeda, Kazuhiro, Kuniko Horie-Inoue, and Satoshi Inoue. 2015. "Identification of Estrogen-Responsive Genes Based on the DNA Binding Properties of Estrogen Receptors Using High-Throughput Sequencing Technology." *Acta Pharmacologica Sinica* 36 (1): 24–31.

Ishii, Kentaro K., and Kazushige Touhara. 2019. “Neural Circuits Regulating Sexual Behaviors via the Olfactory System in Mice.” *Neuroscience Research* 140 (March): 59–76.

Johnson, Zachary V., Manu Tej Sharma Arrojwala, Vineeth Aljapur, Tyrone Lee, Tucker J. Lancaster, Mark C. Lowder, Karen Gu, et al. 2020. “Automated Measurement of Long-Term Bower Behaviors in Lake Malawi Cichlids Using Depth Sensing and Action Recognition.” *Scientific Reports* 10 (1): 20573.

Johnson, Zachary V., Emily C. Moore, Ryan Y. Wong, John R. Godwin, Jeffrey T. Streelman, and Reade B. Roberts. 2020. “Exploratory Behaviour Is Associated with Microhabitat and Evolutionary Radiation in Lake Malawi Cichlids.” *Animal Behaviour* 160 (February): 121–34.

Johnson, Zachary V., Hasse Walum, Yaseen A. Jamal, Yao Xiao, Alaine C. Keebaugh, Kiyoshi Inoue, and Larry J. Young. 2016. “Central Oxytocin Receptors Mediate Mating-Induced Partner Preferences and Enhance Correlated Activation across Forebrain Nuclei in Male Prairie Voles.” *Hormones and Behavior* 79 (March): 8–17.

Johnson, Zachary V., Hasse Walum, Yao Xiao, Paula C. Riefkohl, and Larry J. Young. 2017. “Oxytocin Receptors Modulate a Social Salience Neural Network in Male Prairie Voles.” *Hormones and Behavior* 87 (January): 16–24.

Johnson, Zachary V., and Larry J. Young. 2015. “Neurobiological Mechanisms of Social Attachment and Pair Bonding.” *Current Opinion in Behavioral Sciences* 3 (June): 38–44.

Johnson, Zachary V., and Larry J. Young. 2017. “Oxytocin and Vasopressin Neural Networks: Implications for Social Behavioral Diversity and Translational Neuroscience.” *Neuroscience and Biobehavioral Reviews* 76 (Pt A): 87–98.

Johnson, Zachary V., and Larry J. Young. 2018. “Evolutionary Diversity as a Catalyst for Biological Discovery.” *Integrative Zoology* 13 (6): 616–33.

Jourjine, Nicholas, and Hopi E. Hoekstra. 2021. “Expanding Evolutionary Neuroscience: Insights from Comparing Variation in Behavior.” *Neuron* 109 (7): 1084–99.

Juntti, Scott. 2019. “The Future of Gene-Guided Neuroscience Research in Non-Traditional Model Organisms.” *Brain, Behavior and Evolution* 93 (2–3): 108–21.

Juntti, Scott A., Austin T. Hilliard, Kai R. Kent, Anusha Kumar, Andrew Nguyen, Mariana A. Jimenez, Jasmine L. Loveland, Philippe Mourrain, and Russell D. Fernald. 2016. “A Neural Basis for Control of Cichlid Female Reproductive Behavior by Prostaglandin F2α.” *Current Biology: CB* 26 (7): 943–49.

Jurisch-Yaksi, Nathalie, Emre Yaksi, and Caghan Kizil. 2020. “Radial Glia in the Zebrafish Brain: Functional, Structural, and Physiological Comparison with the Mammalian Glia.” *Glia* 68 (12): 2451–70.

Kastanenka, Ksenia V., Rubén Moreno-Bote, Maurizio De Pittà, Gertrudis Perea, Abel Eraso-Pichot, Roser Masgrau, Kira E. Poskanzer, and Elena Galea. 2020. “A Roadmap to Integrate Astrocytes into Systems Neuroscience.” *Glia* 68 (1): 5–26.

Keifer, Joyce, and Cliff H. Summers. 2016. “Putting the ‘Biology’ Back into ‘Neurobiology’: The Strength of Diversity in Animal Model Systems for Neuroscience Research.” *Frontiers in Systems Neuroscience* 10 (August): 69.

Keleman, Krystyna, Eleftheria Vrontou, Sebastian Krüttner, Jai Y. Yu, Amina Kurtovic-Kozaric, and Barry J. Dickson. 2012. “Dopamine Neurons Modulate Pheromone Responses in Drosophila Courtship Learning.” *Nature* 489 (7414): 145–49.

Kelly, Martin J., and Oline K. Rønnekleiv. 2009. “Control of CNS Neuronal Excitability by Estrogens via Membrane-Initiated Signaling.” *Molecular and Cellular Endocrinology* 308 (1–2): 17–25.

Kimchi, Tali, Jennings Xu, and Catherine Dulac. 2007. “A Functional Circuit Underlying Male Sexual Behaviour in the Female Mouse Brain.” *Nature* 448 (7157): 1009–14.

Kirkpatrick, Mark, and Nick Barton. 2006. “Chromosome Inversions, Local Adaptation and Speciation.” *Genetics* 173 (1): 419–34.

Klinge, C. M. 2001. “Estrogen Receptor Interaction with Estrogen Response Elements.” *Nucleic Acids Research* 29 (14): 2905–19.

Kohl, Johannes, Benedicte M. Babayan, Nimrod D. Rubinstein, Anita E. Autry, Brenda Marin-Rodriguez, Vikrant Kapoor, Kazunari Miyamishi, et al. 2018. “Functional Circuit Architecture Underlying Parental Behaviour.” *Nature* 556 (7701): 326–31.

Küpper, Clemens, Michael Stocks, Judith E. Risse, Natalie Dos Remedios, Lindsay L. Farrell, Susan B. McRae, Tawna C. Morgan, et al. 2016. “A Supergene Determines Highly Divergent Male Reproductive Morphs in the Ruff.” *Nature Genetics* 48 (1): 79–83.

Labusch, Miriam, Laure Mancini, David Morizet, and Laure Bally-Cuif. 2020. “Conserved and Divergent Features of Adult Neurogenesis in Zebrafish.” *Frontiers in Cell and Developmental Biology* 8 (June): 525.

Lacar, Benjamin, Sara B. Linker, Baptiste N. Jaeger, Suguna Krishnaswami, Jerika Barron, Martijn Kelder, Sarah Parylak, et al. 2016. “Nuclear RNA-Seq of Single Neurons Reveals Molecular Signatures of Activation.” *Nature Communications* 7 (April): 11022.

Lai, Cora Sau Wan, Avital Adler, and Wen-Biao Gan. 2018. “Fear Extinction Reverses Dendritic Spine Formation Induced by Fear Conditioning in the Mouse Auditory Cortex.” *Proceedings of the National Academy of Sciences of the United States of America* 115 (37): 9306–11.

Lamichhaney, Sangeet, Guangyi Fan, Fredrik Widemo, Ulrika Gunnarsson, Doreen Schwochow Thalmann, Marc P. Hoeppner, Susanne Kerje, et al. 2016. “Structural Genomic Changes Underlie Alternative Reproductive Strategies in the Ruff (Philomachus Pugnax).” *Nature Genetics* 48 (1): 84–88.

Laurent, Gilles. 2020. “On the Value of Model Diversity in Neuroscience.” *Nature Reviews. Neuroscience* 21 (8): 395–96.

Lévy, Frederic, Martine Batailler, Maryse Meurisse, and Martine Migaud. 2017. “Adult Neurogenesis in Sheep: Characterization and Contribution to Reproduction and Behavior.” *Frontiers in Neuroscience* 11 (October): 570.

Lipsky, Robert H., and Ann M. Marini. 2007. “Brain-Derived Neurotrophic Factor in Neuronal Survival and Behavior-Related Plasticity.” *Annals of the New York Academy of Sciences* 1122 (1): 130–43.

Long, Lijiang, Zachary V. Johnson, Junyu Li, Tucker J. Lancaster, Vineeth Aljapur, Jeffrey T. Streelman, and Patrick T. McGrath. 2020. “Automatic Classification of Cichlid Behaviors Using 3D Convolutional Residual Networks.” *IScience* 23 (10): 101591.

Louilot, A., J. L. Gonzalez-Mora, T. Guadalupe, and M. Mas. 1991. “Sex-Related Olfactory Stimuli Induce a Selective Increase in Dopamine Release in the Nucleus Accumbens of Male Rats. A Voltammetric Study.” *Brain Research* 553 (2): 313–17.

Lukaszewicz, Agnès I., and David J. Anderson. 2011. “Cyclin D1 Promotes Neurogenesis in the Developing Spinal Cord in a Cell Cycle-Independent Manner.” *Proceedings of the National Academy of Sciences of the United States of America* 108 (28): 11632–37.

Lyons, Michelle R., and Anne E. West. 2011. “Mechanisms of Specificity in Neuronal Activity-Regulated Gene Transcription.” *Progress in Neurobiology* 94 (3): 259–95.

Maney, Donna L., Jennifer R. Merritt, Mackenzie R. Prichard, Brent M. Horton, and Soojin V. Yi. 2020. “Inside the Supergene of the Bird with Four Sexes.” *Hormones and Behavior* 126 (104850): 104850.

Martelotto, Luciano G. 2020. "'Frankenstein' Protocol for Nuclei Isolation from Fresh and Frozen Tissue for SnRNAseq Protocol Guidelines." Protocols.Io. December 21, 2020. https://www.protocols.io/view/frankenstein-protocol-for-nuclei-isolation-from-f-5jyl8nx98l2w/v3/guidelines.

Martinelli, David C., Kylie S. Chew, Astrid Rohlmann, Matthew Y. Lum, Susanne Ressl, Samer Hattar, Axel T. Brunger, Markus Missler, and Thomas C. Südhof. 2016. "Expression of C1ql3 in Discrete Neuronal Populations Controls Efferent Synapse Numbers and Diverse Behaviors." *Neuron* 91 (5): 1034–51.

Maruska, Karen P., Russ E. Carpenter, and Russell D. Fernald. 2012. "Characterization of Cell Proliferation throughout the Brain of the African Cichlid Fish Astatotilapia Burtoni and Its Regulation by Social Status." *The Journal of Comparative Neurology* 520 (15): 3471–91.

Maruska, Karen P., and Russell D. Fernald. 2010. "Behavioral and Physiological Plasticity: Rapid Changes during Social Ascent in an African Cichlid Fish." *Hormones and Behavior* 58 (2): 230–40.

Merritt, Jennifer R., Kathleen E. Grogan, Wendy M. Zinzow-Kramer, Dan Sun, Eric A. Ortlund, Soojin V. Yi, and Donna L. Maney. 2020. "A Supergene-Linked Estrogen Receptor Drives Alternative Phenotypes in a Polymorphic Songbird." *Proceedings of the National Academy of Sciences of the United States of America* 117 (35): 21673–80.

Mira, Helena, and Javier Morante. 2020. "Neurogenesis from Embryo to Adult - Lessons from Flies and Mice." *Frontiers in Cell and Developmental Biology* 8 (June): 533.

Moffitt, Jeffrey R., Dhananjay Bambah-Mukku, Stephen W. Eichhorn, Eric Vaughn, Karthik Shekhar, Julio D. Perez, Nimrod D. Rubinstein, et al. 2018. "Molecular, Spatial, and Functional Single-Cell Profiling of the Hypothalamic Preoptic Region." *Science* 362 (6416). https://doi.org/10.1126/science.aau5324.

Moreira, Frederico, Tim-Rasmus Kiehl, Kelvin So, Norbert F. Ajeawung, Carmelita Honculada, Peter Gould, Russell O. Pieper, and Deepak Kamnasaran. 2011. "NPAS3 Demonstrates Features of a Tumor Suppressive Role in Driving the Progression of Astrocytomas." *The American Journal of Pathology* 179 (1): 462–76.

Nagai, Jun, Xinzhu Yu, Thomas Papouin, Eunji Cheong, Marc R. Freeman, Kelly R. Monk, Michael H. Hastings, et al. 2021. "Behaviorally Consequential Astrocytic Regulation of Neural Circuits." *Neuron* 109 (4): 576–96.

Nakazawa, Kazu, Thomas J. McHugh, Matthew A. Wilson, and Susumu Tonegawa. 2004. "NMDA Receptors, Place Cells and Hippocampal Spatial Memory." *Nature Reviews. Neuroscience* 5 (5): 361–72.

Nelson, L. R., and S. E. Bulun. 2001. "Estrogen Production and Action." *Journal of the American Academy of Dermatology* 45 (3 Suppl): S116-24.

Nelson, Sacha B., and Vera Valakh. 2015. "Excitatory/Inhibitory Balance and Circuit Homeostasis in Autism Spectrum Disorders." *Neuron* 87 (4): 684–98.

Newman, S. W. 1999. "The Medial Extended Amygdala in Male Reproductive Behavior. A Node in the Mammalian Social Behavior Network." *Annals of the New York Academy of Sciences* 877 (1 ADVANCING FRO): 242–57.

O'Connell, Lauren A., and Hans A. Hofmann. 2011. "The Vertebrate Mesolimbic Reward System and Social Behavior Network: A Comparative Synthesis." *The Journal of Comparative Neurology* 519 (18): 3599–3639.

O'Connell, Lauren A., and Hans A. Hofmann. 2012. "Evolution of a Vertebrate Social Decision-Making Network." *Science (New York, N.Y.)* 336 (6085): 1154–57.

O'Connell, Lauren A., Bryan J. Matthews, and Hans A. Hofmann. 2012. "Isotocin Regulates Paternal Care in a Monogamous Cichlid Fish." *Hormones and Behavior* 61 (5): 725–33.

Ogawa, S., A. E. Chester, S. C. Hewitt, V. R. Walker, J. A. Gustafsson, O. Smithies, K. S. Korach, and D. W. Pfaff. 2000. "Abolition of Male Sexual Behaviors in Mice Lacking Estrogen Receptors Alpha and Beta (Alpha

Beta ERKO)." *Proceedings of the National Academy of Sciences of the United States of America* 97 (26): 14737–41.

Ogawa, Sonoko, Shinji Tsukahara, Elena Choleris, and Nandini Vasudevan. 2020. "Estrogenic Regulation of Social Behavior and Sexually Dimorphic Brain Formation." *Neuroscience and Biobehavioral Reviews* 110 (March): 46–59.

Patil, Chinar, Jonathan B. Sylvester, Kawther Abdilleh, Michael W. Norsworthy, Karen Pottin, Milan Malinsky, Ryan F. Bloomquist, Zachary V. Johnson, Patrick T. McGrath, and Jeffrey T. Streelman. 2021. "Genome-Enabled Discovery of Evolutionary Divergence in Brains and Behavior." *Scientific Reports* 11 (1): 13016.

Pellegrini, Elisabeth, Nicolas Diotel, Colette Vaillant-Capitaine, Rita Pérez Maria, Marie-Madeleine Gueguen, Ahmed Nasri, Joel Cano Nicolau, and Olivier Kah. 2016. "Steroid Modulation of Neurogenesis: Focus on Radial Glial Cells in Zebrafish." *The Journal of Steroid Biochemistry and Molecular Biology* 160 (June): 27–36.

Pfenning, Andreas R., Erina Hara, Osceola Whitney, Miriam V. Rivas, Rui Wang, Petra L. Roulhac, Jason T. Howard, et al. 2014. "Convergent Transcriptional Specializations in the Brains of Humans and Song-Learning Birds." *Science (New York, N.Y.)* 346 (6215): 1256846.

Purcell, Jessica, Alan Brelsford, Yannick Wurm, Nicolas Perrin, and Michel Chapuisat. 2014. "Convergent Genetic Architecture Underlies Social Organization in Ants." *Current Biology: CB* 24 (22): 2728–32.

Ramallo, Martín R., Agustina Birba, Renato M. Honji, Leonel Morandini, Renata G. Moreira, Gustavo M. Somoza, and Matías Pandolfi. 2015. "A Multidisciplinary Study on Social Status and the Relationship between Inter-Individual Variation in Hormone Levels and Agonistic Behavior in a Neotropical Cichlid Fish." *Hormones and Behavior* 69 (March): 139–51.

Ribbink, A. J., B. A. Marsh, A. C. Marsh, A. C. Ribbink, and B. J. Sharp. 1983. "A Preliminary Survey of the Cichlid Fishes of Rocky Habitats in Lake Malawi." *South African Journal of Zoology* 18 (3): 149–310.

Robinson, Gene E., Russell D. Fernald, and David F. Clayton. 2008. "Genes and Social Behavior." *Science (New York, N.Y.)* 322 (5903): 896–900.

Roesti, Marius, Benjamin Kueng, Dario Moser, and Daniel Berner. 2015. "The Genomics of Ecological Vicariance in Threespine Stickleback Fish." *Nature Communications* 6 (1): 8767.

Salzburger, Walter. 2018. "Understanding Explosive Diversification through Cichlid Fish Genomics." *Nature Reviews. Genetics* 19 (11): 705–17.

Santello, Mirko, Nicolas Toni, and Andrea Volterra. 2019. "Astrocyte Function from Information Processing to Cognition and Cognitive Impairment." *Nature Neuroscience* 22 (2): 154–66.

Sarkar, Saumyendra N., Ren-Qi Huang, Shaun M. Logan, Kun Don Yi, Glenn H. Dillon, and James W. Simpkins. 2008. "Estrogens Directly Potentiate Neuronal L-Type Ca2+ Channels." *Proceedings of the National Academy of Sciences of the United States of America* 105 (39): 15148–53.

Schaal, Sara M., Benjamin C. Haller, and Katie E. Lotterhos. 2022. "Inversion Invasions: When the Genetic Basis of Local Adaptation Is Concentrated within Inversions in the Face of Gene Flow." *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 377 (1856): 20210200.

Schiller, Crystal Edler, Samantha Meltzer-Brody, and David R. Rubinow. 2015. "The Role of Reproductive Hormones in Postpartum Depression." *CNS Spectrums* 20 (1): 48–59.

Shin, Jaehoon, Daniel A. Berg, Yunhua Zhu, Joseph Y. Shin, Juan Song, Michael A. Bonaguidi, Grigori Enikolopov, et al. 2015. "Single-Cell RNA-Seq with Waterfall Reveals Molecular Cascades Underlying Adult Neurogenesis." *Cell Stem Cell* 17 (3): 360–72.

Silva, Vinicius H. da, Veronika N. Laine, Mirte Bosse, Lewis G. Spurgin, Martijn F. L. Derks, Kees van Oers, Bert Dibbits, et al. 2019. “The Genomic Complexity of a Large Inversion in Great Tits.” *Genome Biology and Evolution* 11 (7): 1870–81.

Smith, Caroline C., Lindsey C. Vedder, and Lori L. McMahon. 2009. “Estradiol and the Relationship between Dendritic Spines, NR2B Containing NMDA Receptors, and the Magnitude of Long-Term Potentiation at Hippocampal CA3-CA1 Synapses.” *Psychoneuroendocrinology* 34 Suppl 1 (December): S130-42.

Smith, Chris R., Amy L. Toth, Andrew V. Suarez, and Gene E. Robinson. 2008. “Genetic and Genomic Analyses of the Division of Labour in Insect Societies.” *Nature Reviews. Genetics* 9 (10): 735–48.

Soltesz, Ivan, and Attila Losonczy. 2018. “CA1 Pyramidal Cell Diversity Enabling Parallel Information Processing in the Hippocampus.” *Nature Neuroscience* 21 (4): 484–93.

Srivastava, Deepak P., and Peter Penzes. 2011. “Rapid Estradiol Modulation of Neuronal Connectivity and Its Implications for Disease.” *Frontiers in Endocrinology* 2 (November): 77.

Stefansson, Hreinn, Agnar Helgason, Gudmar Thorleifsson, Valgerdur Steinthorsdottir, Gisli Masson, John Barnard, Adam Baker, et al. 2005. “A Common Inversion under Selection in Europeans.” *Nature Genetics* 37 (2): 129–37.

Stein, Murray B., Chia-Yen Chen, Sonia Jain, Kevin P. Jensen, Feng He, Steven G. Heeringa, Ronald C. Kessler, et al. 2017. “Genetic Risk Variants for Social Anxiety.” *American Journal of Medical Genetics. Part B, Neuropsychiatric Genetics: The Official Publication of the International Society of Psychiatric Genetics* 174 (4): 470–82.

Sylvester, J. B., C. A. Rich, C. Yi, J. N. Peres, C. Houart, and J. T. Streelman. 2013. “Competing Signals Drive Telencephalon Diversity.” *Nature Communications* 4: 1745.

Tilley, F. C., C. Arrondel, C. Chhuon, M. Boisson, N. Cagnard, M. Parisot, G. Menara, et al. 2021. “Disruption of Pathways Regulated by Integrator Complex in Galloway-Mowat Syndrome Due to WDR73 Mutations.” *Scientific Reports* 11 (1): 5388.

Tuttle, Elaina M., Alan O. Bergland, Marisa L. Korody, Michael S. Brewer, Daniel J. Newhouse, Patrick Minx, Maria Stager, et al. 2016. “Divergence and Functional Degradation of a Sex Chromosome-like Supergene.” *Current Biology: CB* 26 (3): 344–50.

Villanueva Espino, Luis Alberto, Adriana Berenice Silva Gómez, and Dolores Adriana Bravo Durán. 2020. “Cognitive Training Increases Dendritic Arborization in the Dorsal Hippocampal CA1 and CA3 Neurons of Female and Male Long-Evans Rats.” *Synapse (New York, N.Y.)* 74 (4): e22140.

Villoutreix, Romain, Diego Ayala, Mathieu Joron, Zachariah Gompert, Jeffrey L. Feder, and Patrik Nosil. 2021. “Inversion Breakpoints and the Evolution of Supergenes.” *Molecular Ecology* 30 (12): 2738–55.

Walton, Clare, Eben Pariser, and Fernando Nottebohm. 2012. “The Zebra Finch Paradox: Song Is Little Changed, but Number of Neurons Doubles.” *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 32 (3): 761–74.

Wilson, Daniel J. 2019. “The Harmonic Mean P-Value for Combining Dependent Tests.” *Proceedings of the National Academy of Sciences of the United States of America* 116 (4): 1195–1200.

Wu, Melody V., Devanand S. Manoli, Eleanor J. Fraser, Jennifer K. Coats, Jessica Tollkuhn, Shin-Ichiro Honda, Nobuhiro Harada, and Nirao M. Shah. 2009. “Estrogen Masculinizes Neural Pathways and Sex-Specific Behaviors.” *Cell* 139 (1): 61–72.

Wu, Ye Emily, Lin Pan, Yanning Zuo, Xinmin Li, and Weizhe Hong. 2017. “Detecting Activated Cell Populations Using Single-Cell RNA-Seq.” *Neuron* 96 (2): 313-329.e6.

Xie, Yuanyuan, and Richard I. Dorsky. 2017. “Development of the Hypothalamus: Conservation, Modification and Innovation.” *Development* 144 (9): 1588–99.

Yang, Eun-Jin, and Walter Wilczynski. 2007. “Social Experience Organizes Parallel Networks in Sensory and Limbic Forebrain.” *Developmental Neurobiology* 67 (3): 285–303.

York, Ryan A., Chinar Patil, Kawther Abdilleh, Zachary V. Johnson, Matthew A. Conte, Martin J. Genner, Patrick T. McGrath, Hunter B. Fraser, Russell D. Fernald, and J. Todd Streelman. 2018. “Behavior-Dependent Cis Regulation Reveals Genes and Pathways Associated with Bower Building in Cichlid Fishes.” *Proceedings of the National Academy of Sciences of the United States of America* 115 (47): E11081–90.

York, Ryan A., Chinar Patil, C. Darrin Hulsey, J. Todd Streelman, and Russell D. Fernald. 2015. “Evolution of Bower Building in Lake Malawi Cichlid Fish: Phylogeny, Morphology, and Behavior.” *Frontiers in Ecology and Evolution* 3 (February). https://doi.org/10.3389/fevo.2015.00018.

Yu, Xinzhu, Anna M. W. Taylor, Jun Nagai, Peyman Golshani, Christopher J. Evans, Giovanni Coppola, and Baljit S. Khakh. 2018. “Reducing Astrocyte Calcium Signaling in Vivo Alters Striatal Microcircuits and Causes Repetitive Behavior.” *Neuron* 99 (6): 1170-1187.e9.

Zhang, Gaoqun, Luisa Lübke, Fushun Chen, Tanja Beil, Masanari Takamiya, Nicolas Diotel, Uwe Strähle, and Sepand Rastegar. 2021. “Neuron-Radial Glial Cell Communication via BMP/Id1 Signaling Is Key to Long-Term Maintenance of the Regenerative Capacity of the Adult Zebrafish Telencephalon.” *Cells (Basel, Switzerland)* 10 (10): 2794.

Zhao, Y., H. Z. Sheng, R. Amini, A. Grinberg, E. Lee, S. Huang, M. Taira, and H. Westphal. 1999. “Control of Hippocampal Morphogenesis and Neuronal Differentiation by the LIM Homeobox Gene Lhx5.” *Science (New York, N.Y.)* 284 (5417): 1155–58.