Core and region enriched networks of behaviorally regulated genes and the singing genome

Osceola Whitney[1†*], Andreas R. Pfenning[1,2†*], Jason T. Howard[1], Charles A Blatti[3], Fang Liu[4], James M.Ward[1], Rui Wang[1], Manolis Kellis[2], Sayan Mukherjee[5], Saurabh Sinha[3], Alexander J. Hartemink[6], Anne E. West[4*], Erich D. Jarvis[1*]

Affiliations:

1. Dept of Neurobiology, Howard Hughes Medical Institute and Duke University Medical Center, Durham, NC 27710 USA
2. Department of Computer Science, Massachusetts Institute of Technology, Cambridge, MA 02139
3. Department of Computer Science, University of Illinois, Urbana-Champaign, IL USA
4. Department of Neurobiology, Duke University Medical Center, Durham, NC 27710 USA
5. Department of Statistics, Duke University, Durham, NC USA
6. Department of Computer Science, Duke University, Durham, NC 27708-0129 USA

* correspondence to: owhitney@gmail.com, apfenning@gmail.com, west@neuro.duke.edu, or jarvis@neuro.duke.edu

† denotes equal contribution

Abstract:

Songbirds represent an important model organism for elucidating molecular mechanisms that link genes with complex behaviors, in part because they have discrete vocal learning circuits that has parallels with those that mediate human speech. We found that ~10% of the genes in the avian genome were regulated by singing, and a striking regional diversity of both basal and singing-induced programs in the four key song nuclei of the zebra finch, a vocal learning songbird. The region enriched patterns were a result of distinct combinations of region-enriched transcription factors (TF), their binding motifs and pre-singing H3K27ac enhancer activity in the regulatory regions of the associated genes. RNAi manipulations validated the role of the calcium-response transcription factor (CaRF) in regulating genes preferentially expressed in specific song nuclei in response to singing. Thus, differential combinatorial binding of a small group of activity-regulated TFs and pre-defined epigenetic enhancer activity influences the anatomical diversity of behaviorally regulated gene networks.

One Sentence Summary:

Singing-induced gene expression arises in the brain through diverse, region-specific gene regulatory networks.

Introduction:

Songbirds offer an important *in vivo* model system for studying transcriptional programs regulated during behavior. This system consists of interconnected brain nuclei that control production of a learned vocal behavior (singing) with parallels to human speech (*1, 2*). Four key song nuclei are embedded within three regionally distinct telencephalic brain cell populations: HVC and LMAN in the nidopallium; RA in the arcopallium; and Area X in the striatum (**Fig. 1A;** (*3-6*)). These nuclei are connected in a vocal motor pathway (HVC to RA) and a vocal learning pathway (LMAN and Area X; (*7-13*)). Human functional analogues to these avian brain regions are in the cortex (pallium) and basal ganglia (striatum) (*2, 6, 14, 15*). This includes song (avian) and speech (human) brain regions that have convergence of differentially expressed genes (*15*) suggesting that the behavioral and neuro-anatomical similarities for the production of learned vocalizations are accompanied by similarities in molecular and genetic mechanisms, such as with FoxP2 (*16*).

The neural activity within song nuclei that underlies singing was initially shown to drive induction of two immediate early genes (IEGs), the transcription factors *EGR1* and *FOS* (*17-19*). Their levels of expression correlate with the amount of singing in a motor-driven and social-context–dependent manner (*20-23*). Subsequent studies identified an additional 33 genes regulated within song nuclei by singing (*24*). The identified gene products have a wide range of cellular and biological process functions (*24*), including from neurogenesis (*25, 26*) to speech (*27, 28*). The genes were also found to cluster in a few anatomical and short temporal patterns of expression, although this was determined manually. As a result we hypothesized that *in vivo* behaviorally induced gene expression may consist of anatomically and temporally diverse gene expression programs that can be regulated by networks of combinatorial transcription factor complexes or epigenetic chromatin differences (*24*). Two reports (*29, 30*) using our oligonucleotide microarrays found many more genes, 800–2,000 gene transcripts, regulated in the song nucleus Area X as a result of singing, but could not test this hypothesis since the data was from only one song nucleus, and/or one time point.

To test this hypothesis, we profiled baseline and singing-regulated gene expression across time in the four key song nuclei using our songbird gene expression microarray, which we annotated based on recently sequenced avian genomes (*15, 31*) and the human genome. Combined with genomic transcription factor motif analyses and chromatin immunoprecipitation sequencing (ChIP-Seq) detection of active chromatin, we find predominantly diverse networks of simultaneously activated cascades of behaviorally regulated genes across brain regions, which can be explained in part by a combination of transcription factor complexes and epigenetic regulatory activity in the genome.

Results:

We analyzed singing-regulated gene expression at a genomic-scale in HVC, LMAN, RA, and Area X of the zebra finch (**Figs. 1, S1**). To do so, we recorded moment-to-moment singing behavior of all animals over a 7-hour time course, laser micro-dissected individual song nuclei from multiple birds at each timepoint, amplified their mRNA, hybridized the resulting cDNA to our custom-designed 44K oligonucleotide microarrays (**Table S1**), and developed a computational approach that yielded a true positive rate >87% as verified by *in situ* hybridization and RT-PCR (**Fig. S2**; **Tables S2, S3**; supplementary material sections [SM] 1-7). This analysis detected 24,498 uniquely expressed transcripts among the four song nuclei in silent and/or singing animals (**Table S4**), of which 18,478 (75%) mapped to 9,059 ENSEMBL v60 annotated

genes of the zebra finch genome, indicating that at least 50% of the transcribed genome is expressed in the song-control circuit of an adult animal during awake behaving hours.

Distinct baseline gene expression profiles define the song circuit

Using a linear model we developed to identify differentially expressed transcripts in each brain region and combinations thereof (SM6), we found that of the 24,498 transcripts, ~5,167 (21%; representing 3,168 genes or approximately 17% of the genes in the avian genome (*29*)) were differentially expressed among song nuclei at baseline in silent animals (i.e. before singing began). These 5,167 transcripts were organized hierarchically into at least 5 major region-specific clusters (**Fig. 2A**; **Table S5**) with different functional enrichments (**Tables S6, S7**). A striatal song nucleus (Area X) cluster was enriched with non-coding RNAs, G-protein coupled receptors and synaptic transmission proteins (**Fig. 2A, turquoise cluster; Table S6**). Cortical-like song nuclei (HVC, LMAN, and RA) were enriched for cell-to-cell signaling membrane-associated, axonal connectivity, and post-synaptic density (PSD) proteins (**Fig. 2A, blue cluster; Table S6**). The nidopallium song nuclei (HVC and LMAN) were further enriched for another group of cell-cell communication and neural connectivity, membrane-associated proteins (**Fig. 2A, yellow cluster; Table S6**). The arcopallium song nucleus RA was enriched for another set of neural connectivity proteins and for proteins involved in epilepsy and Alzheimer's (**Fig. 2A, green cluster; Table S6**). RA was the only pallial brain region that had a large cluster of genes with a lower level of expression, which was enriched for PSD proteins different from the cortical enrichment (**Fig. 2A, brown cluster; Table S6**), and LMAN was the only song nucleus that did not have a large enrichment of genes of its own.

In situ hybridizations of example genes (e.g. some dopamine and glutamate receptors) revealed that most of the song nuclei expression patterns were consistent with the brain subdivisions to which they belonged (**Fig. 3A-C; Table S2**) (*32-34*). However, as seen previously (*33, 35, 36*), some of the song nuclei had highly differential expression from their surrounding brain divisions (i.e. *FMNL1, DGKI, GPSM1* in **Fig. 3A-C**). The most song-nucleus–specific gene was *FAM40B* (aka, *STRIP2*), a phosphatase that was restricted to cortical-like song nuclei and the primary cortical sensory populations (like auditory area L2; **Fig. 3A**).

A dendrogram analysis separated the cortical song nuclei from the striatal, and showed a stronger relationship between HVC and LMAN of the nidopallium (**Figs. 2B, 1A**), consistent with the recently revised understanding of avian brain organization and homologies with mammals (*5, 6, 37*). These findings show that even before singing starts, the song learning nuclei have thousands of differentially expressed genes that define unique molecular functions for each (see (*15*) for characterization of the specializations in song nuclei).

Singing activates both a core and regionally diverse patterns of genes

Of the 24,498 transcripts, we found an estimated 2,740 (~11%) that were singing-regulated, up or down in time, in one or more song nuclei (**Fig. 4A, B**; **Table S8**). These transcripts mapped to 1,833 genes, indicating a conservative estimate of ~10% of the transcribed avian genome that is regulated by singing behavior. Area X had the most regulated transcripts (1,162), followed by HVC (772), RA (702), and LMAN (635; **Fig. 4B**; sum is higher than 2,740 because of transcripts expressed in more than one song nucleus). A small number of genes (82) had singing-regulated splice variant differences (**Table S9**), consistent with splice variant differences at baseline among song nuclei for glutamate receptor subunits (*33*), which can regulate activity-dependent genes in the brain. The vast majority (96%) of the 2,740 singing-regulated transcripts were

enriched in only 1 or 2 song nuclei, and a core set of only about 97 transcripts was regulated in at least three or four (<1.0%) song nuclei; of the latter, only 20 genes were equally regulated in all four song nuclei (**Fig. 4A, B**; **Table S8, green and yellow**).

The core set of 97 transcripts were enriched for known IEGs (*38*), including membrane depolarization-regulated (Ca2+ responsive) genes identified in cultured hippocampal (*39*) and cortical neurons (*40*), and genes induced in the auditory pathway by hearing song (*41*) (**Tables S10A, S7**). In contrast, the brain region-specific singing-regulated genes had very little overlap with classic IEGs or a list of cell cultured-defined depolarization-induced genes (**Table S10A**). Rather, the striatal Area X singing-regulated genes were enriched for cytoskeletal neural connectivity and neural migration functions, and RA was enriched for mitogen-activated protein kinase pathway transcripts, which control gene expression, differentiation, and cell survival. This suggests that our *in vivo* analyses are useful for finding region-specific or stimulus-specific genes that may be relevant for the underlying singing behavior.

Similar to the baseline expression, *in situ* hybridizations revealed that song nuclei expression patterns were consistent with the brain subdivisions to which they belong (**Fig. 3A-C; Table S3**)**,** except that the surrounding brain areas in some birds tended to have lower expression, presumably because they sang without much other movement behavior to cause movement-induced gene expression in the surrounding regions (*42*). We also noted that even among the core early-response genes induced in all song nuclei, expression levels at baseline differed among song nuclei (**Fig. 3D**). This suggests that there is even greater diversity among the song nuclei singing-regulated genes than simply presence or absence of regulation.

Analysis of the behaviorally regulated gene expression across time, using unsupervised hierarchical clustering (SM8), revealed up to 20 temporal profiles (clusters) among the four song nuclei, including transient or sustained, increased or decreased, early (0.5–2 hr) or late (3–7 hr), or two peaks of expression (**Fig. S3A-D; Table S8**). These 20 clusters can be further grouped into four super-clusters of temporal profiles: 1) transient early increases; 2) late-response increases; 3) transient early decreases; and 3) late-response decreases (**Fig. 5A-D**). Only three of the temporal clusters had relatively comparable representations of genes in all brain regions, all belonging to transient early increase clusters, including the IEG 0.5–1 hr cluster **(Figs. 5A, S3, tan cluster; Table S11**), which contained a significant proportion (16%) of the core set of 97 transcripts ($p < 1E-5$**,** hypergeometric test). For the remaining super-temporal profiles, each song nucleus had a region-enriched set of genes, except the late-response increasing pattern in LMAN (**Figs. 5, S3E; Table S11**).

Functional enrichment analyses showed that the activity-regulated gene expression sets from previous cell culture experiments (**Table S7**) were highly enriched in the early transient IEG temporal cluster expressed in all song nuclei (**Table S10B**). All of the late-increase singing-regulated clusters (**Fig. 5B**) also had detectable functional enrichments of genes, with Area X+HVC enriched in calcium ion binding and phosphatase proteins (**blue temporal cluster**); Area X late-increase genes was additionally enriched in chromosome organization, biogenesis (**green**), activity-dependent late-response genes identified in cultured neurons (*40*) (**turquoise**), and ribosomal proteins (**black**); HVC was additionally enriched in RNA-protein complexes and PSD proteins (**cyan**); and RA late-increase genes (**salmon**) was enriched in a different set of calcium ion-binding and ribosomal proteins (**Table S10B; Figs. 5B**). Remarkably, we did not find any functional enrichment for the remaining transiently increased or any of the decreased clusters, except genes regulated by the serum response transcription factor (SRF) in the slow decreasing cluster of RA (**Table S10B; Figs. 5D, yellow**). These findings show that all song

nuclei share a core set of genes with rapid transient up-regulation, but each song nucleus has its own dominant (though partly overlapping) set of other early- and late-responsive behaviorally regulated genes, suggesting cascades of gene regulation specific to each song nucleus with functions that remain to be discovered.

Relationships between differential baseline and differential singing-regulated genes

We next investigated how a small core set of behaviorally regulated transcription factors expressed in most brain regions could regulate a diverse set of downstream genes, with little overlap among regions. We hypothesized that the differential transcriptional state at baseline, before cell stimulation with singing, affects region-enriched singing-regulated expression (*43, 44*). Three lines of evidence support this hypothesis. First, hypergeometric tests revealed significant overlap between subsets of transcripts from the baseline region-enriched clusters (**Fig. 4C, top gray box**) with the singing-regulated region-enriched clusters (**Fig. 4C, red lines; Table S12**) and with 10 of the 20 temporal clusters (**Fig. 4C**, **blue + black lines between two gray boxes**). If a gene was expressed at higher levels in a region relative to others at baseline before singing, it was also more likely to increase in that region during singing; the converse was not true for the decreasing sets of singing-regulated genes.

Second, a genome-wide binding site analysis of motifs for transcription factors (SM11 (*45, 46*)) revealed ~100 motifs enriched in regulatory regions (e.g. directly upstream of transcription start sites) of genes in the temporal behaviorally regulated clusters (**Table S13, S14**; **Fig. 6A**, **B**), and these matched genomic locations also found in mammalian genomes (*47, 48*). With these motifs, we performed an association analysis between the region-specific and temporal clusters of genes to generate song nuclei-specific "transcription factor motif to gene cluster networks" (**Fig. 6C,** simplified network**; Fig. S4,** detailed network**; Table S15,** edge list; statistical significance tested with Euclidean distance to randomly generated networks (SM11-12)). Consistent with the core IEG cluster findings, we found that binding sites for 5 early-activated transcription factors (EATFs: *MEF2*, *SRF*, *NFKB*, *CREB*, and *CaRF*) that are constitutively expressed at baseline and activated in response to neural activity (*38, 49, 50*), were significantly over-represented in the singing-regulated cluster of IEGs expressed in most song nuclei (**Figs. 6C, S4, S5A**). In turn, the binding motifs of the singing-regulated *AP-1* (bound by a FOS-JUN dimer) and *EGR1* IEG transcription factors were also enriched directly upstream of the transcription start sites of many genes in our avian IEG cluster (**Fig. 6A-C**). *EGR1* can bind to its own promoter and down-regulate itself (*51*), which is consistent with the transient increase and subsequent decrease of some transcripts in the IEG temporal cluster. Also over-represented in the IEG cluster was the *ARNT* motif, which also has the binding motif for the IEG *NPAS4*.

Third, consistent with our region-specific clusters, some transcription factors that were differentially expressed in a region or a combination of regions at baseline had binding motifs in genes that were differentially regulated in that region(s) at baseline or during singing. For example, variants of the *NFE2L1* and *MAF* transcription factors that dimerize and bind to the *TCF11* motif (*52*) were higher or lower in Area X relative to the pallial song nuclei at baseline (**Fig. S6**) and the *TCF11* binding motif was over-represented in the slow increase singing-regulated cluster of genes in Area X (**Figs. 6C, S4**, **S5B**). However, there were many other cases where EATFs and other transcription factors did not exhibit differential regional baseline expression but had binding motifs enriched in clusters of singing-regulated genes specific for a song nucleus. For example, the EATF transcription factors *SRF* and *CaRF*, which are not differentially expressed at baseline (**Table S5**), had strong motif associations to singing-

regulated genes in Area X and HVC. The *MZF1* and *PRRX2* transcription factors had associations with different sets of genes in Area X and RA (**Figs. 6C, S4**, **S5B**). Thus, we experimentally tested whether one of these EATFs, *CaRF*, regulated the predicted region-specific genes.

CaRF is required for regulation of both core and regional expressed sets of genes

We investigated the Ca^{2+} responsive transcription factor *CaRF* because the network analyses implicated it in both the regulation of the Ca^{2+} responsive IEGs that are induced in most song nuclei and some that are regionally enriched in Area X and HVC (**Figs. 4C, S6**). Because we lacked an established zebra finch neural cell culture method to test *CaRF* function, we used RNAi against *CaRF* in cultured mouse cortical neurons and hybridized labeled cDNA to mouse oligonucleotide microarrays representing many of the same genes on our zebra finch oligonucleotide microarray (SM4). We identified a set of genes that showed decreased or increased expression after *CaRF* knockdown independent of membrane depolarization (**Fig. S7A; Table S16**), and many of these function in calcium signaling pathways (**Fig. S7; Table S17;** (*53*)). This is consistent with the proposed role of *CaRF* in regulating neuronal gene expression under basal neural activity (*48, 54*), as both a repressor and activator (*48*). Importantly, as predicted by our promoter motif analyses in birds, the ranked list of *CaRF*-regulated genes showed enrichment for singing-regulated genes that had a nearby *CaRF* binding site ($p = 0.0014$ Wilcox test; **Fig. 8B**). This enrichment was highest in the set of genes regulated in Area X and HVC (**Fig. 8B**), supporting our network result (**Fig. 6C**).

CaRF RNAi knockdown also caused genes that were normally up-regulated by membrane depolarization to be suppressed to normal baseline levels and conversely genes that were normally down-regulated by membrane depolarization to be up-regulated (**Fig. 8C; Table S18**). This suggests that *CaRF* is required to buffer activity of these gene promoters under basal conditions such that they can become stimulus-responsive upon membrane depolarization. Importantly, this same set of membrane depolarization- and *CaRF*-regulated genes significantly overlapped with those that had the *CaRF* binding site in the singing-regulated genes of the IEG (tan) cluster. They also significantly overlapped with several other clusters that were specifically upregulated in Area X and HVC (**Fig. 8D, magenta and cyan clusters; Table S19; Fig. S3E**). Genes that showed decreased expression preferentially in RA, but also in other song nuclei (**Fig. S3**, **yellow**) after 2–3 hours of singing (the same amount of time the cultured cells were depolarized) had even greater overlap (**Figs. 8D, yellow**).

Overall, the findings demonstrate a requirement of the *CaRF* transcription factor for baseline and activity-dependent regulation of some of the very same genes for which we found *CaRF* binding motifs that are regulated at baseline and by singing in a region-specific manner, respectively. The calcium signaling and calcium ion binding genes tended to increase during song production and were affected in the *CaRF* knockdown experiments, proving evidence of consistent function across species. We next sought an explanation of how could EATFs that are not differentially expressed at baseline regulate these genes in a region specific manner.

Epigenetic modifications predefine region specificity of gene regulation

Although transcription factors are the ultimate regulators of gene expression, their ability to bind to sites in the genome is gated by chromatin structural changes. Chromatin regulation by acetylation of histone 3 at lysine 27 (H3K27ac) has been extensively studied and shown to be a strong indicator of active enhancers (*55*). We thus performed an experiment to identify active

transcriptional regulatory regions in the genomes of individual dissected song nuclei (RA and Area X, which showed the largest regional differences) before and after singing, as measured by a genome-wide histone ChIP-Seq analysis of H3K27ac (SM14-15; **Table S20**). The active genomic regions can be searched as tracks in the UCSC browser against the zebra finch genome (*56*). This analysis also required that we create a more stringent selection of regional, early- and late singing-responsive genes from the respective clusters in RA and Area X (**Figs. 5, S3**), using principle components analyses (**Fig. S8**).

Out of 35,958 peaks, we found 30% (10,749) enriched in Area X and 21% (7,673) enriched in RA. Under basal conditions, genes with song nuclei-specific expression patterns had nearby genomic regions that were significantly more likely to be marked by H3K27ac in that brain region (**Fig. 8A, blue and red; Table S21;** ~1300 genes). Conversely, genes that were expressed similarly in RA and Area X did not show a significant regional bias in the distribution of this chromatin mark (**Fig. 8A, grey; Table S21**; ~1100 genes examined). Interestingly, when we considered only the set of RA or Area X region-specific genes that were also upregulated by singing, we found that they were already associated with higher nearby H3K27ac in their preferred brain region prior to singing (**Figs. 6B,D,E, S9A-E; Table S22**). There was a strong positive correlation between differences in nearby H3K27ac at baseline and differences in singing dependent upregulation of these genes in RA and Area X (R=0.37, p=1.6E-12; Pearson correlation). Conversely late-response genes that were comparably induced by singing in both RA and Area X showed comparable H3K27ac under basal conditions (**Fig. 8B, grey; Table S22**). Furthermore, the early-response cluster of genes, which were expressed and induced comparably in both RA and Area X (e.g. *FOS*), also showed comparable H3K27ac in both brain regions at baseline (**Figs. 8C, S9A, S10A**). Notably, we did not find any significant difference (e.g. 0 significant peaks, FDR threshold < 0.01) in H3K27ac peaks within either song nucleus when we compared ChIP-Seq profiles obtained before and after singing (**Fig. S10A**). We detected a weak signal for increased H3K27ac peaks in the Area X downregulated genes (**Fig. S10B**).

These data suggest that the regional differences in chromatin activity present before singing begins are predictive of differential singing-dependent induction of late-response genes. This hypothesis was further supported by our observation of regional H3K27ac differences at baseline for 50 genes that had equivalent basal expression in RA and Area X but region-specific upregulation upon singing (**Table S22, blue and red highlights).** An ingenuity pathway analysis on the Area X set of genes out of the 50 mentioned above (**Table S22, blue;** SM15) revealed that they were enriched for locomotion behavior (p=0.004; *ARNTL*, *CALB1*, *FGF14*, *RCAN2*, *RIMS1*) and movement disorder functions (p=0.004; *ARNTL*, *CALB1*, *CAPZB*, *DIRAS2*, *EEF1A2*, *ELMO1*, *FGF14*, *MTMR2*, *RPSA*, *TMED10)* consistent with the function of Area X and the surrounding striatum. There were too few RA-specific genes without baseline differential expression (10 genes) to be tested by pathway analyses. Overall, these findings indicate that region-specific epigenetic chromatin activity at or near transcription factor binding sites for transcription factors expressed in all brain regions could determine which singing or baseline differentially regulated genes are expressed in each brain region.

Discussion:

The magnitude of the anatomical diversity of behaviorally regulated genes and their networks in different brain regions of the same circuit was unexpected (*24, 29, 30, 41*). Our findings suggest this that diversity is controlled by at least two of the following mechanisms: 1) Region-enriched

transcription factors that regulate region-enriched expression of their target genes; and 2) Region-enriched epigenetic marks that determine which genes can be expressed in specific brain regions in both baseline and behaviorally regulated states. The first mechanism is consistent with the hypothesis that interactions between early transcription factors and late response genes coordinates activity-dependent gene induction associated with behavior (*57*), but in this case, in a region specific manner. The second, epigenetic, mechanism is only just beginning to be explored at the level of neural activity (*40, 58*) and has not been addressed in complex behaviors.

Given our findings and known signaling pathways from experiments in cultured cells (*59*) we propose the following overall mechanism (**Abstract Fig. 1**). Neural activity during the performance of a behavior, such as singing, causes release of neurotransmitters at the synapses between connected cells and activates post-synaptic receptors. These receptors initiate an intracellular signaling response that alters the activity, often via phosphorylation, of constitutively expressed EATFs. The activated EATFs bind or are already bound to the open chromatin of promoters or enhancers of the core IEGs enabled in all brain regions, as measured by H3K27ac, to activate their expression. The IEGs in turn, along with EATFs, bind to recognition regions of open chromatin that have already been primed in a cell type-specific manner, which leads to the induction of region-specific late-response genes. Some transcription factors are already expressed in a region-specific manner and add to the diversity of regulation of the downstream genes. Furthermore, our data to show that brain region–specific open enhancers or promoters are already waiting in an active state, ready to do their job at a moment's notice when the neurons fire to turn on programs of gene expression. Thus, the production of learned behavior modulates an already primed transcriptional and epigenetic network specific to different sub-regions of the circuit that controls the behavior.

This model may be an explanation for the finding that the IEG and EATF *NPAS4*, in response to neural activity, activates different sets of genes in cultured excitatory versus inhibitory neurons (*60*). Likewise, we find that common induction of IEGs across the many different kinds of neurons that comprise all song nuclei is associated with distinct programs of late-response genes, which are likely dependent at least in part on IEG regulation. However one notable difference between our data and a recent study of activity-dependent enhancers in cultured neuron preparations is that whereas membrane depolarization was found to further induce H3K27ac at enhancers near activity-regulated genes (*58*), we find that H3K27ac peaks *in vivo* in the brain are already enriched near singing-inducible genes under basal conditions and do not show further activation upon singing. It is possible that the neural networks recruited upon singing are sparse enough in the song nuclei that we were unable to detect H3K27ac changes in these cells against the background noise. An alternative possibility is that ongoing neural activity in the brain of an awake behaving animal is sufficient to keep enhancers poised in a fully active state even prior to execution of a specific behavioral task like singing. In this model, it is regulation of sequence-specific DNA binding of transcription factors that are most important for instructing the level and nature of gene expression, whereas epigenetic marks on chromatin are permissive for expression of the predetermined program.

Our *CaRF* manipulation experiments help reveal further complexity and potential novel mechanisms of activity-dependent gene regulation in the brain. The increased activity-regulated genes that is reversed in the absence of *CaRF* in response to membrane depolarization, suggests that *CaRF* may act as a modulating transcription factor for neural activity–dependent regulation of its target genes. In this scenario, it prevents differential expression of its target genes until neural firing increases. When *CaRF* is removed by knockdown, it can no longer buffer the

expression of these genes in the absence of activity; consequently, in the presence of activity, other factors can regulate the genes in a direction opposite of what *CaR*F would do. The specific mechanisms by which *CaRF* might achieve this function remain to be determined, but the H3K27ac enhancer activity in *CaRF* target genes is likely to play a role.

Additional transcriptional anatomical diversity not tested in this study could possibly be generated with differential expression of neurotransmitter membrane receptors at baseline in different brain regions, which could activate different signaling pathways in those neurons during singing (*2, 33*). Our hypothesis does not explain the down-regulation of some gene clusters where regionally specific transcription factor motifs were not enriched in those genes, and thus their regulation would have to be explained by other mechanisms.

Our findings suggest that each song nucleus has diverse molecular functions and gene networks. Consistent with their dominant roles in song production (*7-13*) compared to other song nuclei, HVC is specifically enriched with singing-regulated increases in PSD proteins used for cell-to-cell communication and RNA-protein complexes and RA is enriched with genes in the MAPK pathway, such as *DUSP1*, which is proposed to be involved in neural protection of a brain region that is highly active during behavior performance (*61, 62*). Consistent with their dominant roles in learning (*7-13*), LMAN shows greater specificity for the CREB pathway, a key transcription factor involved in learning and memory (*59, 63*), and Area X is more enriched with expression of neural connectivity, and chromosome organization and biogenesis genes. In addition, the large over representation of non-coding RNA genes expressed at baseline in Area X indicates that its transcriptional regulatory network may be more extensive than the pallial song nuclei. The larger overrepresentation of neural connectivity and cell signaling genes in the pallial song nuclei indicates greater focus on cell structure and communication.

In terms of memory, a long-held hypothesis is that neural activity will induce an early wave of responsive genes, which in turn regulate a late wave of genes, and that the first wave would act as a molecular switch converting short-term memories into long-term memories (*57, 64, 65*). If true, singing would be associated with continuous memory consolidation and song fine-tuning, with each nucleus having specific waves of gene regulation for their specific functions. An alternative, not mutually exclusive, proposal states that the activity-dependent waves function as a metabolic mechanism to maintain protein turnover for normal cell homeostasis due to increased protein catabolism that occurs during high activity levels (*17*). If true, it would be associated with continued repair of the circuit when used. Our transcription factor binding motif analysis suggests that both the early and late transcriptional responses could be driven by some of the same EATFs. This would indicate that the two waves of gene expression may not entirely depend on each other, and that they could be used for both memory and homeostasis functions.

In summary, as the mechanisms that define the genome-phenotype relationship, including the diversity of gene expression patterns, begins to be understood, so will the role of individual genes and pathways in learning, maintenance, and production of behavior. Performance of complex behavior involves interaction between neural activity, networks of cells, and networks of genes. Untangling the subtle differences in connected neurons, firing patterns, signaling pathways, and transcription factor activity may lead to a greater understanding of the diversity of gene expression patterns we observe here in highly interconnected cells within an intact multicellular organ.

References and Notes:

1. A. J. Doupe, P. K. Kuhl, Birdsong and human speech: common themes and mechanisms. *Annu. Rev. Neurosci.* **22**, 567-631 (1999).
2. E. D. Jarvis, Learned birdsong and the neurobiology of human language. *Annals of the New York Academy of Sciences* **1016**, 749-777 (2004).
3. R. Mooney, M. Konishi, Two distinct inputs to an avian song nucleus activate different glutamate receptor subtypes on individual neurons. *Proc Natl Acad Sci USA* **88**, 4075-4079 (1991).
4. M. Luo, D. J. Perkel, A GABAergic, strongly inhibitory projection to a thalamic nucleus in the zebra finch song system. *J Neurosci* **19**, 6700-6711 (1999).
5. A. Reiner, D. J. Perkel, L. L. Bruce, A. B. Butler, A. Csillag, W. Kuenzel, L. Medina, G. Paxinos, T. Shimizu, G. Striedter, M. Wild, G. F. Ball, S. Durand, O. Güntürkün, D. W. Lee, C. V. Mello, A. Powers, S. A. White, G. Hough, L. Kubikova, T. V. Smulders, K. Wada, J. Dugas-Ford, S. Husband, K. Yamamoto, J. Yu, C. Siang, E. D. Jarvis, O. Gütürkün, A. Forum, Revised nomenclature for avian telencephalon and some related brainstem nuclei. *J Comp Neurol* **473**, 377-414 (2004).
6. E. D. Jarvis, O. Güntürkün, L. Bruce, A. Csillag, H. Karten, W. Kuenzel, L. Medina, G. Paxinos, D. J. Perkel, T. Shimizu, G. Striedter, J. M. Wild, G. F. Ball, J. Dugas-Ford, S. E. Durand, G. E. Hough, S. Husband, L. Kubikova, D. W. Lee, C. V. Mello, A. Powers, C. Siang, T. V. Smulders, K. Wada, S. A. White, K. Yamamoto, J. Yu, A. Reiner, A. B. Butler, A. Consortium, Avian brains and a new understanding of vertebrate brain evolution. *Nat Rev Neurosci* **6**, 151-159 (2005).
7. F. Nottebohm, T. M. Stokes, C. M. Leonard, Central control of song in the canary, Serinus canarius. *J Comp Neurol* **165**, 457-486 (1976).
8. S. W. Bottjer, E. A. Miesner, A. P. Arnold, Forebrain lesions disrupt development but not maintenance of song in passerine birds. *Science* **224**, 901-903 (1984).
9. C. Scharff, F. Nottebohm, A comparative study of the behavioral deficits following lesions of various parts of the zebra finch song system: implications for vocal learning. *J Neurosci* **11**, 2896-2913 (1991)
10. H. Williams, N. Mehta, Changes in adult zebra finch song require a forebrain nucleus that is not necessary for song production. *J Neurobiol* **39**, 14-28 (1999).
11. M. S. Brainard, A. J. Doupe, Interruption of a basal ganglia-forebrain circuit prevents plasticity of learned vocalizations. *Nature* **404**, 762-766 (2000).
12. B. P. Olveczky, A. S. Andalman, M. S. Fee, Vocal experimentation in the juvenile songbird requires a basal ganglia circuit. *PLoS Biol* **3**, e153 (2005).
13. D. Aronov, A. S. Andalman, M. S. Fee, A specialized forebrain circuit for vocal babbling in the juvenile songbird. *Science* **320**, 630-634 (2008).
14. S. W. Bottjer, F. Johnson, Circuits, hormones, and learning: vocal behavior in songbirds. *J Neurobiol* **33**, 602-618 (1997)
15. A. R. Pfenning, E. Hara, O. Whitney, M. R. Rivas, P. Roulhac, J. T. Howard, M. Wirthlin, P. V. Lovell, G. Ganapathy, M. A. Moseley, J. W. Thompson, E. J. Soderblom, C. V. Mello, A. J. Hartemink, E. D. Jarvis, Convergent transcriptional specializations in the brains of humans and song learning birds. *Science*, (in press).
16. C. Scharff, J. Petri, Evo-devo, deep homology and FoxP2: implications for the evolution of speech and language. *Philos Trans R Soc Lond B Biol Sci* **366**, 2124-2140 (2011).

17. E. Jarvis, F. Nottebohm, Motor-driven gene expression. *Proc Natl Acad Sci USA* **94**, 4097-4102 (1997).
18. R. R. Kimpo, A. J. Doupe, FOS is induced by singing in distinct neuronal populations in a motor network. *Neuron* **18**, 315-325 (1997).
19. H. Jin, D. Clayton, Localized changes in immediate-early gene regulation during sensory and motor learning in zebra finches. *Neuron* **19**, 1049-1059 (1997).
20. E. D. Jarvis, C. Scharff, M. R. Grossman, J. A. Ramos, F. Nottebohm, For whom the bird sings: context-dependent gene expression. *Neuron* **21**, 775-788 (1998).
21. O. Whitney, K. Soderstrom, F. Johnson, Post-transcriptional regulation of zenk expression associated with zebra finch vocal development. *Brain Res Mol Brain Res* **80**, 279-290 (2000).
22. O. Whitney, F. Johnson, Motor-induced transcription but sensory-regulated translation of ZENK in socially interactive songbirds. *J. Neurobiol.* **65**, 251-259 (2005).
23. L. Kubikova, E. A. Turner, E. D. Jarvis, The pallial basal ganglia pathway modulates the behaviorally driven gene expression of the motor pathway. *Eur J Neurosci* **25**, 2145-2160 (2007).
24. K. Wada, J. T. Howard, P. McConnell, O. Whitney, T. Lints, M. V. Rivas, H. Horita, M. A. Patterson, S. A. White, C. Scharff, S. Haesler, S. Zhao, H. Sakaguchi, M. Hagiwara, T. Shiraki, T. Hirozane-Kishikawa, P. Skene, Y. Hayashizaki, P. Carninci, E. D. Jarvis, A molecular neuroethological approach for identifying and characterizing a cascade of behaviorally regulated genes. *Pro Natl Aca Sci USA* **103**, 15212-15217 (2006).
25. X. C. Li, E. Jarvis, B. Alvarez-Borda, D. A. Lim, F. Nottebohm, A relationship between behavior, neurotrophin expression, and new neuron survival. *Proc Natl Acad Sci USA* **97**, 8584-8589 (2000).
26. A. J. Lombardino, X.-C. Li, M. Hertel, F. Nottebohm, Replaceable neurons and neurodegenerative disease share depressed UCHL1 levels. *Proc Natl Acad Sci USA* **102**, 8036-8041 (2005).
27. I. Teramitsu, S. A. White, FoxP2 regulation during undirected singing in adult songbirds. *J Neurosci* **26**, 7390-7394 (2006).
28. J. E. Miller, E. Spiteri, M. C. Condro, R. T. Dosumu-Johnson, D. H. Geschwind, S. A. White, Birdsong decreases protein levels of FoxP2, a molecule required for human speech. *J Neurophysiol* **100**, 2015-2025 (2008).
29. W. C. Warren, D. F. Clayton, H. Ellegren, A. P. Arnold, L. W. Hillier, A. Künstner, S. Searle, S. White, A. J. Vilella, S. Fairley, A. Heger, L. Kong, C. P. Ponting, E. D. Jarvis, C. V. Mello, P. Minx, P. Lovell, T. A. F. Velho, M. Ferris, C. N. Balakrishnan, S. Sinha, C. Blatti, S. E. London, Y. Li, Y.-C. Lin, J. George, J. Sweedler, B. Southey, P. Gunaratne, M. Watson, K. Nam, N. Backström, L. Smeds, B. Nabholz, Y. Itoh, O. Whitney, A. R. Pfenning, J. Howard, M. Völker, B. M. Skinner, D. K. Griffin, L. Ye, W. M. McLaren, P. Flicek, V. Quesada, G. Velasco, C. Lopez-Otin, X. S. Puente, T. Olender, D. Lancet, A. F. A. Smit, R. Hubley, M. K. Konkel, J. A. Walker, M. A. Batzer, W. Gu, D. D. Pollock, L. Chen, Z. Cheng, E. E. Eichler, J. Stapley, J. Slate, R. Ekblom, T. Birkhead, T. Burke, D. Burt, C. Scharff, I. Adam, H. Richard, M. Sultan, A. Soldatov, H. Lehrach, S. V. Edwards, S.-P. Yang, X. Li, T. Graves, L. Fulton, J. Nelson, A. Chinwalla, S. Hou, E. R. Mardis, R. K. Wilson, The genome of a songbird. *Nature* **464**, 757-762 (2010).

30. A. T. Hilliard, J. E. Miller, E. R. Fraley, S. Horvath, S. A. White, Molecular microcircuitry underlies functional specification in a basal ganglia circuit dedicated to vocal learning. *Neuron* **73**, 537-552 (2012).
31. G. Zhang. e. al, Comparative genomics reveals insights into avian genome evolution and adaptation. *Science*, (in press).
32. S. Haesler, K. Wada, A. Nshdejan, E. E. Morrisey, T. Lints, E. D. Jarvis, C. Scharff, FoxP2 expression in avian vocal learners and non-learners. *J Neurosci* **24**, 3164-3175 (2004).
33. K. Wada, H. Sakaguchi, E. D. Jarvis, M. Hagiwara, Differential expression of glutamate receptors in avian neural pathways for learned vocalization. *J. Comp. Neurol.* **476**, 44-64 (2004).
34. L. Kubikova, K. Wada, E. D. Jarvis, Dopamine receptors in a songbird brain. *J Comp Neurol* **518**, 741-769 (2010).
35. X. Li, X.-J. Wang, J. Tannenhauser, S. Podell, P. Mukherjee, M. Hertel, J. Biane, S. Masuda, F. Nottebohm, T. Gaasterland, Genomic resources for songbird research and their use in characterizing gene expression during brain development. *Proc Natl Acad Sci USA* **104**, 6834-6839 (2007).
36. P. V. Lovell, D. Clayton, K. L. Replogle, C. V. Mello, Birdsong "transcriptomics": neurochemical specializations of the oscine song system. *PLoS ONE* **3**, e3440 (2008).
37. E. D. Jarvis, J. Yu, M. V. Rivas, H. Horita, G. Feenders, O. Whitney, S. C. Jarvis, E. R. Jarvis, L. Kubikova, A. E. Puck, C. Siang-Bakshi, S. Martin, M. McElroy, E. Hara, J. Howard, A. Pfenning, H. Mouritsen, C. C. Chen, K. Wada, Global view of the functional molecular organization of the avian cerebrum: mirror images and functional columns. *J Comp Neurol* **521**, 3614-3665 (2013).
38. M. R. Lyons, A. E. West, Mechanisms of specificity in neuronal activity-regulated gene transcription. *Prog Neurobiol* **94**, 259-295 (2011).
39. S. W. Flavell, T.-K. Kim, J. M. Gray, D. A. Harmin, M. Hemberg, E. J. Hong, E. Markenscoff-Papadimitriou, D. M. Bear, M. E. Greenberg, Genome-wide analysis of MEF2 transcriptional program reveals synaptic target genes and neuronal activity-dependent polyadenylation site selection. *Neuron* **60**, 1022-1038 (2008).
40. T.-K. Kim, M. Hemberg, J. M. Gray, A. M. Costa, D. M. Bear, J. Wu, D. A. Harmin, M. Laptewicz, K. Barbara-Haley, S. Kuersten, E. Markenscoff-Papadimitriou, D. Kuhl, H. Bito, P. F. Worley, G. Kreiman, M. E. Greenberg, Widespread transcription at neuronal activity-regulated enhancers. *Nature* **465**, 1-10 (2010).
41. S. Dong, K. L. Replogle, L. Hasadsri, B. S. Imai, P. M. Yau, S. Rodriguez-Zas, B. R. Southey, J. V. Sweedler, D. F. Clayton, Discrete molecular states in the brain accompany changing responses to a vocal signal. *Proc Natl Acad Sci USA* **106**, 11364-11369 (2009).
42. G. Feenders, M. Liedvogel, M. Rivas, M. Zapka, H. Horita, E. Hara, K. Wada, H. Mouritsen, E. Jarvis, Molecular mapping of movement-associated areas in the avian brain: a motor theory for vocal learning origin. *PLoS ONE* **3**, e1768 (2008).
43. C. S. Hill, R. Treisman, Growth factors and gene expression: fresh insights from arrays. *Sci STKE* **1999**, PE1 (1999).
44. A. Barco, J. M. Alarcon, E. R. Kandel, Expression of constitutively active CREB protein facilitates the late phase of long-term potentiation by enhancing synaptic capture. *Cell* **108**, 689-703 (2002).

45. S. Sinha, On counting position weight matrix matches in a sequence, with application to discriminative motif finding. *Bioinformatics* **22**, e454-e463 (2006).
46. J. Kim, R. Cunningham, B. James, S. Wyder, J. D. Gibson, O. Niehuis, E. M. Zdobnov, H. M. Robertson, G. E. Robinson, J. H. Werren, S. Sinha, Functional characterization of transcription factor motifs using cross-species comparison across large evolutionary distances. *PLoS computational biology* **6**, e1000652 (2010).
47. A. R. Pfenning, R. Schwartz, A. L. Barth, A comparative genomics approach to identifying the plasticity transcriptome. *BMC Neurosci* **8**, 20 (2007).
48. A. R. Pfenning, T.-K. Kim, J. M. Spotts, M. Hemberg, D. Su, A. E. West, Genome-wide identification of calcium-response factor (CaRF) binding sites predicts a role in regulation of neuronal signaling pathways. *PLoS ONE* **5**, e10870 (2010).
49. S. W. Flavell, M. E. Greenberg, Signaling Mechanisms Linking Neuronal Activity to Gene Expression and Plasticity of the Nervous System. *Annu. Rev. Neurosci.* **31**, 563-590 (2008).
50. S. Loebrich, E. Nedivi, The function of activity-regulated genes in the nervous system. *Physiol Rev* **89**, 1079-1103 (2009).
51. X. Cao, R. Mahendran, G. Guy, Detection and characterization of cellular EGR-1 binding to its recognition site. *J Biol Chem* **268**, 16949-16957 (1993).
52. O. O. Johnsen, N. N. Skammelsrud, L. L. Luna, M. M. Nishizawa, H. H. Prydz, A. B. A. Kolstø, Small Maf proteins interact with the human transcription factor TCF11/Nrf1/LCR-F1. *Nucleic Acids Res* **24**, 4289-4297 (1996).
53. A. Subramanian, P. Tamayo, V. K. Mootha, S. Mukherjee, B. L. Ebert, M. A. Gillette, A. Paulovich, S. L. Pomeroy, T. R. Golub, E. S. Lander, J. P. Mesirov, Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA* **102**, 15545-15550 (2005).
54. K. A. McDowell, A. N. Hutchinson, S. J. E. Wong-Goodrich, M. M. Presby, D. Su, R. M. Rodriguiz, K. C. Law, C. L. Williams, W. C. Wetsel, A. E. West, Reduced Cortical BDNF Expression and Aberrant Memory in Carf Knock-Out Mice. *J Neurosci* **30**, 7453-7465 (2010).
55. E. P. Consortium, An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57-74 (2012).
56. https://genome.ucsc.edu/cgi-bin/hgTracks?hgS_doOtherUser=submit&hgS_otherUserName=apfenning&hgS_otherUserSessionName=taeGut1_basic. *UCSC Genome Browser*, (2014).
57. P. Goelet, V. F. Castellucci, S. Schacher, E. R. Kandel, The long and the short of long-term memory--a molecular framework. *Nature* **322**, 419-422 (1986).
58. A. N. Malik, T. Vierbuchen, M. Hemberg, A. A. Rubin, E. Ling, C. H. Couch, H. Stroud, I. Spiegel, K. K. Farh, D. A. Harmin, M. E. Greenberg, Genome-wide identification and characterization of functional neuronal activity-dependent enhancers. *Nat Neurosci* **17**, 1330-1339 (2014).
59. A. E. West, M. E. Greenberg, Neuronal Activity-Regulated Gene Transcription in Synapse Development and Cognitive Function. *Cold Spring Harb Perspect Biol* **3**, a005744-a005744 (2011).
60. I. Spiegel, A. R. Mardinly, H. W. Gabel, J. E. Bazinet, C. H. Couch, C. P. Tzeng, D. A. Harmin, M. E. Greenberg, Npas4 regulates excitatory-inhibitory balance within neural circuits through cell-type-specific gene programs. *Cell* **157**, 1216-1229 (2014).

61. H. Horita, K. Wada, M. V. Rivas, E. Hara, E. D. Jarvis, The dusp1 immediate early gene is regulated by natural stimuli predominantly in sensory input neurons. *J Comp Neurol* **518**, 2873-2901 (2010).
62. H. Horita, M. Kobayashi, W.-c. Liu, K. Oka, E. D. Jarvis, K. Wada, Specialized Motor-Driven dusp1 Expression in the Song Systems of Multiple Lineages of Vocal Learning Birds. *PLoS ONE* **7**, e42173 (2012).
63. K. Sakamoto, K. Karelina, K. Obrietan, CREB: a multifaceted regulator of neuronal plasticity and protection. *J Neurochem* **116**, 1-9 (2011).
64. J. I. Morgan, T. Curran, Stimulus-transcription coupling in neurons: role of cellular immediate-early genes. *TINS* **12**, 459-462 (1989).
65. C. H. Bailey, D. Bartsch, E. R. Kandel, Toward a molecular definition of long-term memory storage. *Proc Natl Acad Sci USA* **93**, 13445-13452 (1996).
66. J. M. Wettenhall, G. K. Smyth, limmaGUI: a graphical user interface for linear modeling of microarray data. *Bioinformatics* **20**, 3705-3706 (2004).
67. P. L. Roulhac, J. M. Ward, J. W. Thompson, E. J. Soderblom, M. Silva, M. A. Moseley, E. D. Jarvis, Microproteomics: Quantitative Proteomic Profiling of Small Numbers of Laser-Captured Cells. *Cold Spring Harbor Protocols* **2011**, 218-234 (2011).
68. K. Replogle, A. P. Arnold, G. F. Ball, M. Band, S. Bensch, E. A. Brenowitz, S. Dong, J. Drnevich, M. Ferris, J. M. George, G. Gong, D. Hasselquist, A. G. Hernandez, R. Kim, H. A. Lewin, L. Liu, P. V. Lovell, C. V. Mello, S. Naurin, S. Rodriguez-Zas, J. Thimmapuram, J. Wade, D. F. Clayton, The Songbird Neurogenomics (SoNG) Initiative: community-based tools and strategies for study of brain gene function and evolution. *BMC Genomics* **9**, 131 (2008).
69. B. J. Haas, S. L. Salzberg, W. Zhu, M. Pertea, J. E. Allen, J. Orvis, O. White, C. R. Buell, J. R. Wortman, Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. *Genome Biol* **9**, R7 (2008).
70. P. Flicek, M. R. Amode, D. Barrell, K. Beal, S. Brent, Y. Chen, P. Clapham, G. Coates, S. Fairley, S. Fitzgerald, L. Gordon, M. Hendrix, T. Hourlier, N. Johnson, A. Kähäri, D. Keefe, S. Keenan, R. Kinsella, F. Kokocinski, E. Kulesha, P. Larsson, I. Longden, W. McLaren, B. Overduin, B. Pritchard, H. S. Riat, D. Rios, G. R. S. Ritchie, M. Ruffier, M. Schuster, D. Sobral, G. Spudich, Y. A. Tang, S. Trevanion, J. Vandrovcova, A. J. Vilella, S. White, S. P. Wilder, A. Zadissa, J. Zamora, B. L. Aken, E. Birney, F. Cunningham, I. Dunham, R. Durbin, X. M. Fernández-Suarez, J. Herrero, T. J. P. Hubbard, A. Parker, G. Proctor, J. Vogel, S. M. J. Searle, Ensembl 2011. *Nucleic Acids Res* **39**, D800-806 (2011).
71. A. R. Quinlan, I. M. Hall, BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-842 (2010).
72. K. D. Pruitt, T. Tatusova, D. R. Maglott, NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res* **35**, D61-65 (2007).
73. D. A. Benson, I. Karsch-Mizrachi, D. J. Lipman, J. Ostell, E. W. Sayers, GenBank. *Nucleic Acids Res* **37**, D26-31 (2009); published online EpubJan 1 (10.1093/nar/gkn723).
74. R. C. Gentleman, V. J. Carey, D. M. Bates, B. Bolstad, M. Dettling, S. Dudoit, B. Ellis, L. Gautier, Y. Ge, J. Gentry, K. Hornik, T. Hothorn, W. Huber, S. Iacus, R. Irizarry, F. Leisch, C. Li, M. Maechler, A. J. Rossini, G. Sawitzki, C. Smith, G. Smyth, L. Tierney, J.

Y. H. Yang, J. Zhang, Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol* **5**, R80 (2004).
75. W. Stacklies, H. Redestig, M. Scholz, D. Walther, J. Selbig, pcaMethods--a bioconductor package providing PCA methods for incomplete data. *Bioinformatics* **23**, 1164-1167 (2007).
76. W. Huber, A. von Heydebreck, H. Sültmann, A. Poustka, M. Vingron, Variance stabilization applied to microarray data calibration and to the quantification of differential expression. *Bioinformatics* **18 Suppl 1**, S96-104 (2002).
77. J. M. George, H. Jin, W. S. Woods, D. F. Clayton, Characterization of a novel protein regulated during the critical period for song learning in the zebra finch. *Neuron* **15**, 361-372 (1995).
78. J. H. Ward, Hierarchical Grouping to Optimize an Objective Function. *J Amer Statistical Association* **58**, 236-244 (1963).
79. K. S. Pollard, S. Dudoit, M. J. Van Der Laan, Multiple testing procedures: R multtest package and applications to genomics. http://biostats.bepress.com/ucbbiostat/paper164/, (2004).
80. A. S. Adler, S. Sinha, T. L. A. Kawahara, J. Y. Zhang, E. Segal, H. Y. Chang, Motif module map reveals enforcement of aging by continual NF- B activity. *Genes Dev* **21**, 3244-3257 (2007).
81. S. Sinha, A. S. Adler, Y. Field, H. Y. Chang, E. Segal, Systematic functional characterization of cis-regulatory motifs in human core promoters. *Genome Res* **18**, 477-488 (2008).
82. C. Alaux, S. Sinha, L. Hasadsri, G. J. Hunt, E. Guzmán-Novoa, G. DeGrandi-Hoffman, J. L. Uribe-Rubio, B. R. Southey, S. Rodriguez-Zas, G. E. Robinson, Honey bee aggression supports a link between gene regulation and behavioral evolution. *Proc Natl Acad Sci USA* **106**, 15400-15405 (2009).
83. S. Sinha, X. Ling, C. W. Whitfield, C. Zhai, G. E. Robinson, Genome scan for cis-regulatory DNA motifs associated with social behavior in honey bees. *Proc Natl Acad Sci USA* **103**, 16352-16357 (2006).
84. J. C. Bryne, E. Valen, M.-H. E. Tang, T. Marstrand, O. Winther, I. da Piedade, A. Krogh, B. Lenhard, A. Sandelin, JASPAR, the open access database of transcription factor-binding profiles: new content and tools in the 2008 update. *Nucleic Acids Res* **36**, D102-106 (2008).
85. E. Wingender, The TRANSFAC project as an example of framework technology that supports the analysis of genomic regulation. *Brief Bioinformatics* **9**, 326-332 (2008).
86. J. D. Storey, R. Tibshirani, Statistical significance for genomewide studies. *Proc Natl Acad Sci USA* **100**, 9440-9445 (2003).
87. M. Kazemian, Q. Zhu, M. S. Halfon, S. Sinha, Improved accuracy of supervised CRM discovery with interpolated Markov models and cross-species comparison. *Nucleic Acids Res* **39**, 9463-9472 (2011).
88. P. Zhou, M. Porcionatto, M. Pilapil, Y. Chen, Y. Choi, K. F. Tolias, J. B. Bikoff, E. J. Hong, M. E. Greenberg, R. A. Segal, Polarized signaling endosomes coordinate BDNF-induced chemotaxis of cerebellar precursors. *Neuron* **55**, 53-68 (2007).
89. R. A. Irizarry, B. M. Bolstad, F. Collin, L. M. Cope, B. Hobbs, T. P. Speed, Summaries of Affymetrix GeneChip probe level data. *Nucleic Acids Res* **31**, e15 (2003).

90. T. Beissbarth, T. P. Speed, GOstat: find statistically overrepresented Gene Ontologies within a group of genes. *Bioinformatics* **20**, 1464-1465 (2004).
91. M. Hollander, D. A. Wolfe, *Nonparametric statistical methods*. (John Wiley and Sons, New York, ed. 2nd, 1999), pp. 787.
92. J. Michaud, K. M. Simpson, R. Escher, K. Buchet-Poyau, T. Beissbarth, C. Carmichael, M. E. Ritchie, F. Schutz, P. Cannon, M. Liu, X. Shen, Y. Ito, W. H. Raskind, M. S. Horwitz, M. Osato, D. R. Turner, T. P. Speed, M. Kavallaris, G. K. Smyth, H. S. Scott, Integrative analysis of RUNX1 downstream pathways and target genes. *BMC Genomics* **9**, 363 (2008).
93. H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, S. Genome Project Data Processing, The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079 (2009).
94. T. Bailey, P. Krajewski, I. Ladunga, C. Lefebvre, Q. Li, T. Liu, P. Madrigal, C. Taslim, J. Zhang, Practical guidelines for the comprehensive analysis of ChIP-seq data. *PLoS computational biology* **9**, e1003326 (2013).
95. P. V. Kharchenko, M. Y. Tolstorukov, P. J. Park, Design and analysis of ChIP-seq experiments for DNA-binding proteins. *Nat Biotechnol* **26**, 1351-1359 (2008).
96. A. T. Lun, G. K. Smyth, De novo detection of differentially bound regions for ChIP-seq data using peaks and windows: controlling error rates correctly. *Nucleic Acids Res* **42**, e95 (2014).
97. Q. Li, J. B. Brown, H. Huang, P. J. Bickel, Measuring reproducibility of high-throughput experiments. *Ann. Appl. Stat* **5**, 1752-1779 (2011).
98. E. P. Consortium, E. Birney, et al, Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* **447**, 799-816 (2007).
99. M. Love, W. Hunber, S. Anders, Moderated estimation of fold change and dispersion for RNA-Seq data with DESeq2. *bioRxiv*, (2014)http://dx.doi.org/10.1101/002832).
100. J. M. Wettenhall, G. K. Smyth, limmaGUI: a graphical user interface for linear modeling of microarray data. *Bioinformatics* **20**, 3705-3706 (2004).
101. E. Garcia-Calero, C. Scharff, Calbindin expression in developing striatum of zebra finches and its relation to the formation of area X. *J Comp Neurol* **521**, 326-341 (2013).
102. T. A. F. Velho, P. Lovell, C. V. Mello, Enriched expression and developmental regulation of the middle-weight neurofilament (NF-m) gene in song control nuclei of the zebra finch. *J. Comp. Neurol.* **500**, 477-497 (2006).
103. T. K. Kim, M. Hemberg, J. M. Gray, A. M. Costa, D. M. Bear, J. Wu, D. A. Harmin, M. Laptewicz, K. Barbara-Haley, S. Kuersten, E. Markenscoff-Papadimitriou, D. Kuhl, H. Bito, P. F. Worley, G. Kreiman, M. E. Greenberg, Widespread transcription at neuronal activity-regulated enhancers. *Nature* **465**, 182-187 (2010).
104. W. C. Warren, D. F. Clayton, H. Ellegren, A. P. Arnold, L. W. Hillier, A. Kunstner, S. Searle, S. White, A. J. Vilella, S. Fairley, A. Heger, L. Kong, C. P. Ponting, E. D. Jarvis, C. V. Mello, P. Minx, P. Lovell, T. A. Velho, M. Ferris, C. N. Balakrishnan, S. Sinha, C. Blatti, S. E. London, Y. Li, Y. C. Lin, J. George, J. Sweedler, B. Southey, P. Gunaratne, M. Watson, K. Nam, N. Backstrom, L. Smeds, B. Nabholz, Y. Itoh, O. Whitney, A. R. Pfenning, J. Howard, M. Volker, B. M. Skinner, D. K. Griffin, L. Ye, W. M. McLaren, P. Flicek, V. Quesada, G. Velasco, C. Lopez-Otin, X. S. Puente, T. Olender, D. Lancet, A. F. Smit, R. Hubley, M. K. Konkel, J. A. Walker, M. A. Batzer, W. Gu, D. D. Pollock, L. Chen, Z. Cheng, E. E. Eichler, J. Stapley, J. Slate, R. Ekblom, T. Birkhead, T. Burke, D.

Burt, C. Scharff, I. Adam, H. Richard, M. Sultan, A. Soldatov, H. Lehrach, S. V. Edwards, S. P. Yang, X. Li, T. Graves, L. Fulton, J. Nelson, A. Chinwalla, S. Hou, E. R. Mardis, R. K. Wilson, The genome of a songbird. *Nature* **464**, 757-762 (2010).

105. R. Schippert, F. Schaeffel, M. P. Feldkaemper, Microarray analysis of retinal gene expression in Egr-1 knockout mice. *Molecular vision* **15**, 2720-2739 (2009).

106. T. Aid-Pavlidis, P. Pavlidis, T. Timmusk, Meta-coexpression conservation analysis of microarray data: a "subset" approach provides insight into brain-derived neurotrophic factor regulation. *BMC Genomics* **10**, 420 (2009).

107. J. Boros, I. J. Donaldson, A. O'Donnell, Z. A. Odrowaz, L. Zeef, M. Lupien, C. A. Meyer, X. S. Liu, M. Brown, A. D. Sharrocks, Elucidation of the ELK1 target gene network reveals a role in the coordinate regulation of core components of the gene regulation machinery. *Genome Res* **19**, 1963-1973 (2009).

108. S. Dong, K. L. Replogle, L. Hasadsri, B. S. Imai, P. M. Yau, S. Rodriguez-Zas, B. R. Southey, J. V. Sweedler, D. F. Clayton, Discrete molecular states in the brain accompany changing responses to a vocal signal. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 11364-11369 (2009).

109. S. W. Flavell, T. K. Kim, J. M. Gray, D. A. Harmin, M. Hemberg, E. J. Hong, E. Markenscoff-Papadimitriou, D. M. Bear, M. E. Greenberg, Genome-wide analysis of MEF2 transcriptional program reveals synaptic target genes and neuronal activity-dependent polyadenylation site selection. *Neuron* **60**, 1022-1038 (2008).

110. J. Lalonde, P. E. Lachance, A. Chaudhuri, Developmental and activity-dependent genomic occupancy profiles of CREB in monkey area V1. *Genes Brain Behav* **8**, 149-160 (2009).

111. M. D. Croning, M. C. Marshall, P. McLaren, J. D. Armstrong, S. G. Grant, G2Cdb: the Genes to Cognition database. *Nucleic Acids Res* **37**, D846-851 (2009).

112. R. P. Haberman, H. J. Lee, C. Colantuoni, M. T. Koh, M. Gallagher, Rapid encoding of new information alters the profile of plasticity-related mRNA transcripts in the hippocampal CA3 region. *Proc Natl Acad Sci USA* **105**, 10601-10606 (2008).

113. S. C. Vernes, E. Spiteri, J. Nicod, M. Groszer, J. M. Taylor, K. E. Davies, D. H. Geschwind, S. E. Fisher, High-throughput analysis of promoter occupancy reveals direct neural targets of FOXP2, a gene mutated in speech and language disorders. *Am J Hum Genet* **81**, 1232-1250 (2007).

114. E. Spiteri, G. Konopka, G. Coppola, J. Bomar, M. Oldham, J. Ou, S. C. Vernes, S. E. Fisher, B. Ren, D. H. Geschwind, Identification of the transcriptional targets of FOXP2, a gene linked to speech and language, in developing human brain. *Am J Hum Genet* **81**, 1144-1157 (2007).

115. A. E. McKee, N. Neretti, L. E. Carvalho, C. A. Meyer, E. A. Fox, A. S. Brodsky, P. A. Silver, Exon expression profiling reveals stimulus-mediated exon use in neural cells. *Genome Biol* **8**, R159 (2007).

116. G. Xiang, L. Pan, W. Xing, L. Zhang, L. Huang, J. Yu, R. Zhang, J. Wu, J. Cheng, Y. Zhou, Identification of activity-dependent gene expression profiles reveals specific subsets of genes induced by different routes of Ca(2+) entry in cultured rat cortical neurons. *Journal of cellular physiology* **212**, 126-136 (2007).

117. K. Sugino, C. M. Hempel, M. N. Miller, A. M. Hattox, P. Shapiro, C. Wu, Z. J. Huang, S. B. Nelson, Molecular taxonomy of major neuronal classes in the adult mouse forebrain. *Nat Neurosci* **9**, 99-107 (2006).

Acknowledgments:
We thank S. Augustin, J. Coleman and N. Diaz Nelson for help collecting and processing singing bird samples for microarray analysis, Dr. H. Dressman and L.-L. Rowlette at the Duke Institute for Genome Science & Policy for microarray hybridizations, and J-N. Audet for song nuclei dissections for the H3K27ac experiments. We thank A. Cavanaugh, M. Lawson, and M. Dean for useful comments after careful reading of an earlier version of this manuscript. OW was supported by post-doctoral training grants from the APA and NSF 0610337. This work was supported by grants from the National Institutes of Deafness and Communication Disorders grant R01DC007218 and the Howard Hughes Medical Institute to EDJ. Microarray data have been submitted to the Gene Expression Omnibus (http://www.ncbi.nlm.nih.gov/geo) under accession no. GSE33365. Annotations are also available in **Table S1**. Raw files of the Chip-Seq H3K27ac experiments are at (http://www.broadinstitute.org/~pfenning/finchWig/).

Figure Legends:

Fig. 1. Song system and laser micro-dissection. (**A**) Sagittal schematic of the zebra finch brain showing positions and some connections of song nuclei. Pallial, striatal, and pallidal regions are distinguished by colors. Black arrows, posterior vocal pathway involved in song production; white arrows, anterior vocal pathway involved in song learning and modulation; dashed arrows, connections between the two pathways. Smaller song nuclei of NIF, Av, and MO are not shown. (**B**) Song nuclei were laser-capture microdissected from males that were either silent or continuously singing for 0.5 hr, 1 hr, and for each hour thereafter, up to 7 hr, resulting in over 200 total microarrays. Shown are images of 10 µm tissue sections before and after laser capture microdissection at 10X magnification. Before: Following dehydration, song nuclei fiber density appears darker than surrounding tissue. After: Song nuclei regions are selectively cut out using an infrared laser. Capture: The cut song nuclei transferred to the cap by the LCM system. For microarray analysis each of the 4 song nuclei from each animal was captured separately to individual LCM caps. Dorsal is up, anterior is right. Scale bar 2 mm.

Fig. 2. Region-enriched gene expression at baseline. (**A**) A heatmap of hierarchically clustered expression profiles of 5,167 transcripts (rows) that are differentially expressed across regions at baseline (FDR $q < 0.1$; see **Fig. S11** for FDR $q < 0.2$) in silent birds (red: increases; blue: decreases; white: no change) relative to mean Area X expression (numbers of transcripts not shown for small size clusters). Each transcript is normalized to the average value of expression in Area X. Each column is an animal replicate. Detailed results are in **Table S4**. (**B**) Average linkage hierarchical tree, generated from mean expression in each brain region, representing the molecular expression relationships between regions.

Fig. 3 *In situ* hybridizations of baseline and singing-regulated genes. (**A**) Genes higher in all pallial song nuclei (RA, HVC, LMAN) relative to the striatal song nucleus (Area X) at baseline (**Fig. 2A,** blue clusters). (**B**) Genes differentially expressed just among the pallial song nuclei (green, yellow, and brown clusters) at baseline. (**C**) Genes higher in the striatal song nucleus relative to pallial song nuclei (turquoise cluster). (**D**) Core singing-regulated genes regulated in 3–4 song nuclei detected by microarrays, but detected in all 4 with diverse levels by *in situ* hybridization, most peaking at 30 minutes. (**E**) Region-enriched singing-regulated genes in one

or two song nuclei, with peaks of expression at later time points. Film autoradiograph images are inverted showing white as labeled mRNA expression of the gene indicated below the image. Dorsal is up, anterior is right. Scale bar 2 mm.

Fig. 4. Region-enriched gene expression in response to singing. (**A**) A 4-way Venn diagram showing regional singing-regulated distribution of 2,740 transcripts (FDR $q < 0.2$). (**B**) Heatmap all 2,740 transcripts from the Venn diagram, hierarchically clustered independently in all four song nuclei, and then sorted by increased or decreased expression, and level of significance from highest to lowest in the linear model. Each column (170 total) is an animal replicate within a time point, and white lines separate time points. Red: increases; blue: decreases; white: no change relative to 0-hr samples for each song nucleus. Each transcript is normalized so that the maximum increase relative to non-singing birds in any region is the darkest shade of red for increasing transcripts and the maximum decrease is the darkest shade of blue for decreasing transcripts. Boxes highlight significant behaviorally regulated enrichment for each region (FDR $q < 0.2$ for that region). **Fig. S12** shows a more stringent heat map of region-enriched expression with a similar result. (**C**) Relationships among clusters of transcripts from the baseline region-enriched (top grey box, from **Fig. 2A**), singing temporal-enriched (rectangular nodes, from **Fig. S3A–D**), and singing region-enriched (bottom grey box, from **panel B**) patterns. Nodes are colored according to their cluster colors in the respective figures. Edges between two nodes correspond to significant overlap between two groups of transcripts ($p < 0.001$, hypergeometric test). Nodes are sorted to optimize non-crossing of edges. Detailed results are in Table S8.

Fig. 5. Temporal singing-regulated patterns across time. (**A**) Averages of gene expression levels in four temporal clusters of transient early response increases. (**B**) Averages of six late-response gene cluster increases. (**C**) Averages of four transient early response cluster decreases. (**D**) Averages of six late-response gene cluster decreases. The temporal profiles are normalized such that non-singing birds have a value of 0 and each gene has a maximum increase or decrease of 1. Each point represents the mean across all gene-brain region combinations for that time point. The 20 colors match the major temporal clusters in **Fig. S3A-D**.

Fig. 6. Transcription factor binding motifs found in singing-regulated genes. (**A**) Location bias of the target window of several motifs relative to its nearby gene when the motif search was confined to the local promoter, i.e., 5 kb upstream and 2 kb downstream of the start of the first nucleotide of the first exon of the gene. Fold change (plotted on the log scale y-axes) is the ratio of the percentage of the motif target windows that fell within a particular position category relative to the first exon of a gene (target %) versus the percentage of windows that fall within that position category genome-wide (genome %). (**B**) Location bias of the motif target window relative to its nearby gene when the motif search was performed over the gene territory, i.e., half way upstream and half way downstream to the last or first exon of the nearest, non-overlapping gene. (**C**) Transcription factor motif-gene cluster network summarized from **Fig. S4** showing relationships between enriched EATFs (gray circles) and their binding motifs in subsets of genes from the temporal singing-regulated clusters (colored rectangular nodes as in **Fig. S3A–D**). Edges are colored on the basis of the region specific expression of the predicted regulatory targets of the TF within each singing-regulated cluster (SM11-12). Detailed results are in Table S13 and Figure S4.

Fig. 7. RNAi knockdown illuminates *CaRF* binding motif relationships with singing-regulated genes. (**A**) Heatmap of genes affected by *CaRF* knockdown independent of membrane depolarization in mouse cultured neurons. Rows represent the 100 transcripts most changed by *CaRF* RNAi knockdown ($p < 0.0014$, FDR $q < 0.475$), sorted according to the t-statistic, which takes direction of regulation into account. Each column is an independent sample ($n = 3$ unstimulated controls, $n = 3$ KCl depolarized in the presence of either scrambled RNAi or *CaRF* RNAi knockdown virus). Color intensities (blue to red) represent the log fold change in knockdown cells relative to the mean of the scrambled control conditions. (**B**) Significance of the enrichment of zebra finch baseline genes (cluster colors according to **Fig. 2A**) with *CaRF* promoter motifs in the ranked list of t-values for *CaRF* knockdown affected genes in mouse cultured neurons. $p < 0.05$ (above line) is a significant association, Wilcox rank sum statistic over multiple permutations (*66*). (**C**) Similar to (A), except for genes that respond differently to KCl activity in the *CaRF* knockdown cells. Rows represent the 100 transcripts most changed in expression ($p < 0.015$, factorial test), sorted according to the t-statistic. (**D**) Significance of the enrichment of zebra finch singing-regulated genes (cluster colors according to **Figs. 5, S3**) with *CaRF* promoter motifs in the ranked list of t-values for genes differentially regulated by neural activity in mouse cortical neurons during *CaRF* knockdown versus control. $p < 0.05$ (above line) is a significant association, Wilcox rank sum statistic over multiple permutations (*66*).

Fig. 8. Region-specific epigenetic signatures predefine behaviorally regulated gene expression. (**A**) Density plot of genes differentially expressed at baseline in RA vs. Area X and the difference in the level of nearby H3K27ac peaks in the genomes of cells in RA X vs. Area X. Each H3K27ac peak is mapped to a gene with the nearest transcription start site. For each gene, the change in all mapped H3K27ac peaks are averaged. The H3K27ac distributions for RA vs. Area X enriched genes are significantly different (p=1.5E-186, t test). (**B**) Similar plot as in (A) except for differentially expressed late-response singing-regulated genes. The distributions for RA and Area X are also significantly different (p=1.8E-5, t test). Note, however, there are two peaks in RA, which suggest active genomic sites in Area X in the negative peak for RA could be genes that are actively suppressed in Area X. Corresponding data can be found in **Tables S21**, S22. (**C**) H3K27ac peaks surrounding a gene induced by singing in across all brain regions, FOS; (**D**) H3K27ac peaks of a gene induced specifically in Area X, PTPN5. (**E**) H3K27ac peaks of a gene induced at low levels in RA but not detectable in Area X, BDNF. The plots show the log-likelihood ratios of H3K27ac signal in pooled baseline RA and pooled baseline Area X samples versus input DNA around the genomic regions in the zebra finch. The relevant gene models from the UCSC genome browser are shown below. Peaks measure both enhancer and promoter regions. Left of the H3K27ac peaks are in-situ hybridization mRNA signal in singing animals. FOS and PTPN5 are shown in Fig. 3 and BDNF is used with permission from (*37*).