

Vocal learning-associated convergent evolution in mammalian proteins and regulatory elements

Morgan E. Wirthlin^{1†‡§}, Tobias A. Schmid^{2†}, Julie E. Elie^{2,3†}, Xiaomeng Zhang^{1¶}, Amanda Kowalczyk^{1‡}, Ruby Redlich¹, Varvara A. Shvareva⁴, Ashley Rakuljic⁴, Maria B. Ji⁵, Ninad S. Bhat⁴, Irene M. Kaplow^{1‡}, Daniel E. Schäffer^{1#}, Alyssa J. Lawler^{6‡¶}, Andrew Z. Wang¹, BaDoi N. Phan^{1‡}, Siddharth Annaldasula¹, Ashley R. Brown^{1‡}, Tianyu Lu¹, Byung Kook Lim⁷, Eiman Azim⁸, Zoonomia Consortium^{**}, Nathan L. Clark⁹, Wynn K. Meyer¹⁰, Sergei L. Kosakovsky Pond¹¹, Maria Chikina¹², Michael M. Yartsev^{2,3*††}, Andreas R. Pfenning^{1*‡††}

¹Department of Computational Biology, Carnegie Mellon University, Pittsburgh, PA 15213, USA. ²Helen Wills Neuroscience Institute, University of California, Berkeley, Berkeley, CA 94708, USA. ³Department of Bioengineering, University of California, Berkeley, Berkeley, CA 94708, USA. ⁴Department of Molecular and Cell Biology, University of California, Berkeley, Berkeley, CA 94708, USA. ⁵Department of Psychology, University of California, Berkeley, Berkeley, CA 94708, USA. ⁶Department of Biological Sciences, Carnegie Mellon University, Pittsburgh, PA 15213, USA. ⁷Neurobiology section, Division of Biological Science, University of California, San Diego, La Jolla, CA 92093, USA. ⁸Molecular Neurobiology Laboratory, Salk Institute for Biological Studies, La Jolla, CA 92037, USA. ⁹Department of Biological Sciences, University of Pittsburgh, Pittsburgh, PA 15213, USA. ¹⁰Department of Biological Sciences, Lehigh University, Bethlehem, PA 18015, USA. ¹¹Department Of Biology, Temple University, Philadelphia, PA 19122, USA. ¹²Department of Computational and Systems Biology, University of Pittsburgh School of Medicine, Pittsburgh, PA 15213, USA. †These authors contributed equally to this work. ‡Present address: Present address: Department of Biomedical Engineering, Duke University, Durham, NC 27705. §Present address: Allen Institute for Brain Science, Seattle, WA 98109, USA. ¶Present address: Broad Institute, Massachusetts Institute of Technology, Cambridge, MA 02142, USA. #Present address: Massachusetts Institute of Technology, Cambridge, MA 02139. **Zoonomia Consortium collaborators and affiliations are listed at the end of this paper. ††These authors jointly supervised this work.

*Corresponding author. Email: myartsev@berkeley.edu (M.M.Y.); apfenning@cmu.edu (A.R.P.)

Vocal production learning is a convergently evolved trait in vertebrates. To identify brain genomic elements associated with mammalian vocal learning, we integrated genomic, anatomical and neurophysiological data from the Egyptian fruit-bat with analyses of the genomes of 215 placental mammals. First, we identified a set of proteins evolving more slowly in vocal learners. Then, we discovered a vocal-motor cortical region in the Egyptian fruit-bat, an emergent vocal learner, and leveraged that knowledge to identify active *cis*-regulatory elements in the motor cortex of vocal learners. Machine learning methods applied to motor cortex open chromatin revealed 50 enhancers robustly associated with vocal learning whose activity tended to be lower in vocal learners. Our research implicates convergent losses of motor cortex regulatory elements in mammalian vocal learning evolution.

Vocal production learning—the ability of an organism to modify the acoustic properties of its vocalizations as a result of social experience—is an example of convergent evolution, having evolved independently within multiple lineages of birds and mammals, including humans, where it manifests as speech (Fig. 1A) (1, 2). Vocal production learning (“vocal learning”) has been extensively studied in songbirds, highlighting numerous shared behavioral features of birdsong and speech learning, such as a dependence on auditory input during a critical developmental period and a juvenile babbling phase of sensorimotor exploration prior to the maturation of the adult vocalizations (1). Convergence between song-learning birds and humans extends to neuroanatomical specializations, including direct corticobulbar projections from the vocal motor cortex analog to the hindbrain motoneurons controlling the vocal apparatus (3) and shared transcriptional specializations in analogous speech- and song-specialized brain regions (4). Thus, songbirds have become a premier model for exploring the fundamental brain anatomical, molecular, and genomic features

associated with vocal learning (1, 3). An expanding literature on vocal learning behavior across mammals suggests an underappreciated diversity in the phenotypic expression of vocal learning across the taxa traditionally thought to possess it (2, 5–8). Study of the diverse forms of mammalian vocal learning behaviors could broaden our understanding of the core molecular, anatomical and physiological brain mechanisms of vocal learning and of the mechanisms underlying the convergent evolution of skilled motor behaviors more broadly.

We evaluated evidence of convergent genomic specializations shared among four lineages of vocal learning mammals that independently evolved this trait — bats, cetaceans, pinnipeds, and humans (Fig. 1A) — using whole-genome datasets and recently developed computational approaches. Specifically, we used protein-coding sequences from genomes generated by the Zoonomia Consortium (9, 10) and models of evolutionary rate convergence (11, 12) to identify protein-coding regions repeatedly associated with the evolution of vocal learning across mammals. Although

we found 200 protein-coding genes significantly associated with vocal learning, none of them showed strong evidence of selection in all four mammalian clades and only five showed strong evidence of selection in three out of the four clades. Due to individual lineages contributing disproportionately to many of the protein-coding results, we hypothesized that non-coding regulatory elements might also be under constraint for the evolution of vocal learning. We next profiled open chromatin, a proxy for regulatory element activity (13), in multiple brain regions and somatic tissues in the Egyptian fruit bat, a mammal with robust vocal plasticity (14–16) to identify vocalization-associated regulatory genomic specializations. We accomplished this by combining anatomical tracing and electrophysiological recordings in vocalizing bats to identify a region of motor cortex associated with vocal production. The vocalization-associated epigenomic data collected from this region of this bat species — combined with hundreds of mammalian genomes (17, 18), their associated reference-free whole-genome alignments (19), and high-quality epigenomic data from the motor cortex of multiple additional mammalian species (20–22) — provided the foundation to apply a machine learning approach, the Tissue-Aware Conservation Inference Toolkit (TACIT) (23). This approach allowed us to identify putative enhancers, distal regulatory elements that tend to be highly tissue-specific, associated with the convergent evolution of vocal learning. In sum, we combined recently developed computational tools and neuroanatomical experiments in the Egyptian fruit bat and found evidence of convergent evolution in both protein-coding and noncoding DNA sequences.

Results

Convergent evolution in protein sequence associated with vocal learning behavior

To explore the possibility of shared genomic specializations associated with vocal learning, we first applied RERconverge (11) to recently released protein-coding alignments obtained for hundreds of mammals (10) to identify protein-coding genes whose relative rates of evolution differ between vocal learners and other mammals, and which may thus be under selection related to vocal learning (11, 24). We analyzed 16,209 high-quality protein-coding gene alignments across 215 boreoeutherian mammals, including 26 vocal learning species, 164 vocal non-learners, and 25 species without confident annotations (Fig. 1A and data S1) (Materials and Methods). We found evidence for lower evolutionary rates in vocal learners compared to non-learners in 804 genes and evidence for elevated evolutionary rates in 102 genes (Tau adj. $p < 0.01$ and permutations (24) adj. $p < 0.01$) (Data S2 and fig. S1). To identify which specific clades were driving the differential rates of evolution, we applied a Bayes Factor analysis that examines each clade individually for evidence of selection (18) (data S2). Despite the large number of significant associations based on RERconverge, we found no single protein-coding gene with consistently

lower or elevated evolutionary rates in all four vocal learning clades. Among the genes with reduced evolutionary rates in vocal learning species, we found only five out of 804 protein-coding genes with strong evidence of selective pressure in three out of the four vocal learning clades: *CENPC*, *CATSPERG*, *MGA*, *TREML2*, and *ZCWPW1* (Bayes Factor > 5) (Fig. 1, B and C). None of these proteins reached the threshold for selection in the human lineage (Materials and Methods), which could indicate different mechanisms of evolution in the laurasiatherian vocal learning clades relative to humans. Our results suggest that the vast majority of protein-coding genes that we identify are evolving much faster or much slower in one clade, but are only weakly associated with vocal learning across the other lineages.

The most strongly associated genes were *CENPC* (Fig. S2A; Tau = -0.30; Tau adj. $p = 8.7 \times 10^{-6}$; permutations adj. $p \leq 0.001$), and *GRM8* (Fig. S2B, Tau = 0.26; Tau adj. $p = 3.5 \times 10^{-4}$; permutations adj. $p \leq 0.001$). *GRM8* represents an especially promising candidate because it has previously been linked to anatomical specializations for vocal learning in songbirds (25) and is a known target of the speech-associated *FOXP2* transcription factor (26). In primates and rodents, both *GRM8* and *FOXP2* are markers of a rare class of medium spiny neurons implicated in motor control that coexpress both *DRD1* and *DRD2* dopamine receptors (27–29). Overall, our results support a model where vocal learning behavior is only partially explained by differences in protein-coding gene evolutionary rates.

To further explore selection on these vocal learning-associated protein-coding genes, we applied an additional set of tools from the HyPhy package that compares non synonymous (dN) and synonymous (dS) substitution rates in the nucleotides of the amino acid sequence (30). Here, these sensitive evolutionary models of codon substitution formally compare selective regimes, modeled as dN/dS distributions, between branches annotated with the vocal learning phenotype and the rest of the phylogenetic tree (12, 30). The results were largely consistent with amino-acid level methods: the protein-coding genes with lower protein evolutionary rates in vocal learning clades also tended to be under higher constraint in vocal learning clades (Wilcoxon $p = 2.5 \times 10^{-4}$), and the genes with higher evolutionary rates in vocal learning species showed evidence of accelerated evolution in these same clades (Wilcoxon $p = 7.9 \times 10^{-9}$) (fig. S3 and data S3). To further explore the functional trends of protein-coding genes associated with the evolution of vocal learning, we focused on the set of proteins that showed consistent behavior between RERconverge and the HyPhy RELAX model (Benjamini-Hochberg false discovery rate $q < 0.05$). This yielded a set of 126 proteins that were more slowly evolving in vocal learning clades and 74 with an elevated rate of evolutionary changes relative to other species. The complementary approaches of RERconverge and HyPhy RELAX identify a total of 200 vocal learning-associated genes and suggest that this behavior, vocal learning, is having a

substantial impact on protein evolutionary rates.

We further interrogated evolutionary pressures across the vocal learning-associated genes by looking for evidence of diversifying position selection using the HyPhy BUSTED-PH model. Evidence of diversifying positive selection was found in 6.3% (13 transcripts, 9 genes) within the set of genes with elevated rates of evolution in vocal learning species based on RER-Converge and the HyPhy RELAX model. As expected, we identified much lower rates of diversifying positive selection in the gene with lower rates of evolution in vocal learning species (1.0%) and within the set of randomly chosen transcripts (2.0%). Among the 9 genes that showed evidence of positive selection, 8 have been associated with neurodevelopment (*CCDC136*, *KIDINS220*, *LRRN1*, *RSG5*, *CYLD*, *GABRA5*, *NETO2*, *KIAA1109*) (31–38). The gene *CCDC136* has more directly been associated with multiple language-related phenotypes in humans (31, 39, 40). These results suggest that the vocal learning-associated genes across mammals may tend to play a role in human brain development and vocal behavior.

To more systematically explore the functional trends within these 126 and 74 protein-coding genes, we conducted a gene ontology analysis using EnrichR. Protein-coding genes with lower evolutionary rates in vocal learning species were associated with *Regulation of DNA-templated Transcription* ($p=1.10 \times 10^{-6}$, adj. $p=1.9 \times 10^{-5}$), *Regulation of Canonical Wnt Signaling Pathway* ($p=5.1 \times 10^{-5}$, adj. $p=0.013$), and the *Autism* human phenotype ontology ($p=5.8 \times 10^{-6}$, adj. $p=0.0028$) (Fig. 1, F and G, and data S4). The genes with accelerated evolutionary rates were not enriched for any pathways at an adjusted $p < 0.05$ with at least 5 genes contributing. The enrichment of autism-associated genes among the set of genes with greater levels of conservation in vocal learners (*MECP2*, *RAD21*, *DYRK1A*, *SIM1*, *FTSJ1*, *MEIS2*, *FGFR1*) is particularly interesting given prevalence of speech delay and early vocal production differences in human subjects with autism (41, 42) and the previous association between autism genes and the evolution of vocal behavior in birds and bats (43, 44). Based on the association with autism, we further explored the function of the vocal learning-conserved gene set in the context of early vocal production differences. Although only four human loci have been associated with differences in early vocal production, protein-coding genes overlapping two of these loci show higher levels of conservation in vocal learning clades (*INSC*, *DAPK3*) (45).

Identification of a vocal production region in egyptian fruit bats

The enrichment of transcription factors in the set of vocal learning-associated proteins suggests that differences in gene regulation are likely to be a major factor in the evolution of vocal learning. Since gene regulation is often tissue-specific, we sought to identify motor regions of the brain involved in vocal production and contrast their epigenomic profiles with motor regions not involved in vocal production. We conducted this comparison in the

Egyptian fruit bat, *Rousettus aegyptiacus*, a mammalian species with robust vocal plasticity (16, 46) and with data on its motor cortex mapping (47). To identify a candidate region, we were guided by the hypothesis that fine vocal-motor control, a key ability to vocal learning, may be associated with the anatomical specialization of the motor cortex (48–51). In particular, previous work suggested that a cortical region controlling complex vocal behavior would be characterized by a direct, monosynaptic projection onto the motoneurons controlling the vocal source (in mammals, the larynx) (48–52). Such a direct connection has been observed robustly in humans (53–56) and vocal learning birds [songbirds, parrots and hummingbirds, (57–59)], but has not been reliably found in vocal non-learning species such as chimpanzees (41) or mice (60).

We first determined whether a direct corticobulbar anatomical connection existed in *R. aegyptiacus*. Guided by cortical mapping experiments (47), we injected anterograde tracers into the part of the motor cortex that has been associated with orofacial motor control (ofM1) and identified fluorescently labeled descending cortical fibers in the hindbrain region where the laryngeal motoneurons reside: the nucleus ambiguus (NA) (Fig. 2A; fig. S4, A and B; and movie S1). To test the existence of a direct monosynaptic projection, we also specifically identified laryngeal motoneurons in the NA by retrogradely labeling them through bilateral muscular injection of CTB (Cholera Toxin B) into the cricothyroid muscles of the bat larynx (Fig. 2A). We validated the colocalization of descending cortical fibers and local synaptic boutons with laryngeal motoneurons using two complementary labeling approaches: one relying on immunostaining of synapses (VGLUT1) and the other one using viral labeling of synapses (SYN) (Fig. 2, B to F, and fig. S5). Across five bats, 79.2% of the retrogradely labeled motoneurons (61/77) colocalized with descending cortical fibers and 26% of them (20/77) colocalized with both cortical fibers and synaptic boutons, pointing to the existence of a robust direct corticospinal projection to laryngeal motoneurons (Fig. 2G). This colocalization in the NA was consistent across the different techniques (Fig. 2G and fig. S5) and could not be found in any other brainstem motor nuclei, including the hypoglossal nucleus, which controls the tongue and neck muscles (fig. S4, C to E). We noted that the corticobulbar fibers crossed the midline anterior to the NA at the level of the facial nucleus, offering a direct contralateral path for the innervation of the NA (fig. S4F). These anatomical findings highlight the bat ofM1 as a possible candidate region associated with vocal production.

To further corroborate the role of ofM1 in vocal control, we tested whether ongoing single-cell neural activity in this area was associated with vocal production. We performed wireless electrophysiological recordings from four bats engaged in free vocal interactions with peers (Fig. 2H). Vocalizations were identified and recorded using wireless call detectors placed around the necks of the individual bats (see Materials and Methods, (46)). We found

that about half of the recorded single units in ofM1 (115/237) showed a significant change in firing rates when the bats produced vocalizations as compared to staying quiet (fig. S6, A to C) (ANOVA with a Poisson Generalized Linear Model per cell; *p*-value threshold = 0.001; Materials and Methods). In 25% of ofM1 cells that were excited during vocal production (26/104), the change of activity could not be accounted for by jaw or tongue movements, indicating that these cells were engaged in the motor control of movements specific to vocal production (fig. S6D). Furthermore, many of the single units had a sustained increase of activity during production of vocalizations, but not during perception of vocalizations (Fig. 2I). To further assess this specific neural modulation during vocal-motor production, we quantified the information between the time-varying firing rate and the amplitude modulation of the vocalizations. This analysis confirmed that ofM1 neurons had significantly higher motor than auditory information (Fig. 2J) (likelihood-ratio test on LME models, $N=219$, $LRStat = 62.515$, $df = 1$, $p = 2.6645 \times 10^{-5}$; average d-prime change in information gain during motor production = 0.15 ± 0.13 , corresponding to an increase of 0.286 ± 0.035 bits/s). Combined, the results of the anatomical and electrophysiological study defined ofM1 as a motor cortical area associated with vocal production in *R. aegyptiacus*.

Epigenomic specializations in the vocal production region of the egyptian fruit bat motor cortex

We next sought to epigenomically profile candidate regulatory elements in vocal and non-vocal brain regions in *R. aegyptiacus* to identify vocal learning-associated regulatory genomic specializations. We generated a multi-tissue atlas of open chromatin data — indicative of regulatory activity — by performing ATAC-seq (assay for transposase-accessible chromatin sequencing (61)) across 7 brain regions and 3 somatic tissues of *R. aegyptiacus* (Materials and Methods), including ofM1 (Fig. 3A). From a total set of 88,389 noncoding, non-promoter open chromatin regions (OCRs) in primary motor cortex (M1), we identified 348 candidate enhancers with differential open chromatin between orofacial motor cortex (ofM1) and wing motor cortex (wM1) ($p < 0.05$) (Fig. 3B and data S5) (Materials and Methods). Genes proximal to OCRs with differential open chromatin between ofM1 and wM1 were significantly enriched for functional association with neuronal projections and transcriptional regulation (Data S6). These included OCRs near the genes of 51 known transcription factors (TFs), including *FOXP2*, a TF that has been extensively implicated in human speech and vocal learning (Fig. 3C) (62). Notably, genes near OCRs differentially open between bat ofM1 and wM1 included genes we had identified as being under convergent acceleration in vocal learners using evolutionary approaches: RERconverge analysis ($n = 11$) and the HyPhy RELAX analysis ($n=3$; *GATA3*, *LRRN1*, *TNIP3*) (data S6). These specialized regions of open chromatin, coupled with an enrichment of transcription

factors in the set of vocal learning-associated protein-coding genes, suggest that both *cis* and *trans* differences in gene regulation contribute to the evolution of vocal learning behaviors.

Convergent evolution in candidate enhancer sequences associated with vocal learning behavior

Since there is accumulating evidence that *cis*-regulatory differences in enhancer regions are driving the evolution of complex traits (63–65), we sought to identify OCRs whose tissue- and cell type-specificity would be shared across species of vocal learners. Detecting *cis*-regulatory element differences associated with trait evolution is challenging because many enhancers can preserve the same regulatory function even when the underlying genome sequence is highly divergent, and many *cis*-regulatory elements have tissue-specific activity (66–68). Thus, methods for convergent evolution that rely on the alignment of individual nucleotides between species (11, 69, 70) are likely to miss a substantial proportion of key candidate enhancers.

We therefore sought to extend our search for *cis*-regulatory elements whose evolution is associated with vocal learning behavior using a recently developed machine learning approach, TACIT [Tissue-Aware Conservation Inference Toolkit (23)]. Given that it is infeasible to map the brains and collect motor cortex tissue from each vocal learning and closely related non-learning species, the TACIT approach uses machine learning models (23) to predict motor cortex open chromatin across orthologous regions of the genome (66–68). TACIT then associates predictions with vocal learning in a way that corrects for phylogenetic relationships (Fig. 4A). We used the predictions from convolutional neural networks (CNNs) that were previously trained using DNA sequence-based M1 open chromatin data obtained in this study for *R. aegyptiacus* with ATAC-seq and collected earlier for the mouse (21), the rat, and the macaque (20) to predict motor cortex open chromatin across 222 mammalian genomes (Materials and Methods) (23). Given that parvalbumin has been shown to be a shared marker of brain areas critical for vocal learning in songbirds and humans (4), we also used CNNs trained to predict cell type-specific open chromatin using ATAC-seq data from mouse and human M1 parvalbumin-positive neurons (M1-PV+) (22, 23, 71). We identified regions whose predicted open chromatin was consistently lower or higher in vocal learners relative to vocal non-learners using phylogenetic logistic regression (72, 73) with phylogenetic permutations (24) [permutations adj. $p < 0.1$; Materials and Methods, (23)] (Fig. 4A). We identified 33 open chromatin regions from our M1 CNN models that had lower predicted open chromatin in vocal learning species and 11 that had higher predicted open chromatin in vocal learning species (Fig. 4D, table S1, and data S7). From the M1-PV+ predictions, we identified five candidate enhancers that had lower predicted activity in vocal learning species and one candidate enhancer that had higher predicted regulatory activity in vocal learning species. (table S2 and

data S7). Unlike the protein-coding genes, the majority of vocal learning-associated enhancers showed evidence of higher or lower activity in at least three out of the four vocal learning clades (Fig. 4, B to D, and data S8). Consistent with the finding that convergent vocal learning-associated gene regulation is primarily repressive (4), we found that the majority of candidate enhancers ($n = 38/50$ OCRs, 76%) had lower predicted open chromatin activity in vocal learning relative to vocal non-learning mammals (fig. S7).

To interpret potential functions of the vocal learning-associated candidate enhancers, we annotated the nearest genes in the mouse (data S9). In many cases, the genes closest to these putative enhancers have been associated with significant developmental delay or complete absence of speech when disrupted in humans (tables S1 and S2). Four of the OCRs identified by the M1 model were proximal to genes—*GALC*, *TCF4*, *TSHZ3*, and *ZNF536*—that were also near OCRs with differential activity between bat ofM1 and wM1 (data S6). Two of the vocal learning-associated M1 OCRs were proximal to genes—*DAAM1* and *VIP*—previously shown to have convergent gene expression between humans and song-learning birds (4). To further explore the function of the vocal learning-associated OCRs in the motor cortex, we annotated their cell type-specificity using publicly available mouse BICCN data (74). The cell type most enriched for OCRs predicted to be lower in vocal learning species was Layer V ET (extratelencephalic) neurons, which have previously been implicated in vocal learning (75), but strong enrichments were also found for other cortical excitatory neurons (Fig. 4, E and F, and data S10).

Among genes near vocal learning-associated OCRs, the *DACT1* (TACIT adj. $p = 0.0014$; RERconverge Tau adj. $p < 0.0001$) and *CELF4* (TACIT adj. $p < 0.023$; RERconverge Tau adj. $p < 0.0034$) proteins also displayed significantly lower relative evolutionary rates in vocal learners. Despite the lack of direct evidence in the literature for its role in speech production, *CELF4* has been associated with autism in the human population (76) and its function in Layer V pyramidal neurons has been linked with seizures in mice (77).

Multiple M1-PV+ interneuron OCRs associated with vocal learning are near genes previously associated with autism. For example, an OCR that is negatively associated with vocal learning evolution is in an intron of the gene *CCSER1*, which has nonsense mutations implicated in autism (78) and is in a locus associated with musical beat synchronization (79). An OCR that is positively associated with the evolution of vocal learning is in an intron of the gene *CNTNAP4*, whose deletions and copy number variation in humans and mice have been implicated in neurological disorders, including autism in humans (80, 81). To test whether these associations would have been identified by chance, we tested whether vocal learning-associated OCRs tended to be near genes associated with autism. We found the M1-PV+ OCRs with human orthologs near genes associated with autism (82) tend to be more significantly associated with vocal learning evolution than other

OCRs with human orthologs (Wilcoxon $p = 0.0071$).

Discussion

Convergent evolution of vocal production learning has been associated with convergent evolution at the neuroanatomical level: cortical motor regions driving vocal production in humans and songbirds (human motor cortex and songbird RA) show increased connectivity with the brainstem and striatum (3). These same motor regions also show convergent evolution in patterns of gene expression, with commonly decreased gene expression found in both song-learning birds and humans (4). In this study, we investigated convergent evolution of vocal learning in mammals, both at the anatomical and the genetic level. First, we found a direct motor corticobulbar connection from a cortical region implicated in vocal production in a vocal learning bat. Second, we revealed widespread evidence of convergent evolution across vocal learning mammals in protein-coding sequences and candidate regulatory enhancers.

Our parallel study of both coding and noncoding regions linked with the vocal learning trait identifies many protein-coding genes (200) and a smaller number of noncoding regions (50), distal sites of open chromatin, that are associated with vocal learning. Although a larger number of significant protein-coding genes are identified, the vast majority of these are primarily driven by strong evidence in one of the vocal learning clades and only weak evidence in the other three. In contrast, the majority of significant noncoding regions show robust evidence of convergent selective pressure in at least three out of the four clades. The larger number of identified proteins relative to open chromatin regions could be due to better statistical power from being able to directly model nucleotide evolution in protein-coding sequence, which tend to be more stable than regulatory elements across species (11, 12). We note, however, that only 5/200 of the significantly associated protein-coding genes showed robust evidence of differential rates of evolution in at least three of four vocal learning clades. Out of these 200 proteins, many were neurodevelopmental transcription factors, which are among the most highly conserved genes in mammals (83) and thus likely to play roles in a broad range of contexts that could constrain their evolution. It is also possible that some of the identified proteins could be associated with other convergent traits that correlate with vocal learning across mammals, including echolocation (bats, cetaceans), marine adaptations (cetaceans, pinnipeds), or increased longevity (bats, cetaceans, humans).

In contrast, 33/50 vocal learning-associated OCRs had differential predicted open chromatin in at least three of four vocal learning clades; this independent convergence of gene regulatory function suggests that these OCRs may be critical for the evolution of vocal learning. Enhancers tend to have functions that are much more context- and tissue-specific (68), making them less functionally constrained than protein-coding genes, which could

perhaps allow more flexibility for an individual enhancer to evolve a new role for a specific trait like vocal learning. In sum, our results suggest that the evolution of mammalian vocal learning is largely driven by changes to the noncoding regulatory elements that orchestrate gene expression rather than to the protein-coding genes themselves.

Despite the different methodologies applied to identify convergent evolution in coding and noncoding regions, both protein coding- and regulatory element-focused approaches implicated gene functional pathways associated with human autism. In our protein-coding analyses, genes with lower evolutionary rates based on both RERconverge and HyPhy RELAX were enriched for autism function. Likewise, in our analyses of regulatory evolution, multiple autism-linked genes were near human orthologs of the vocal learning-associated M1-PV OCRs. In humans, autism is often associated with speech delays and differences in social behavior, both of which could be related to the evolutionary trait of vocal learning ability (84). Broadly, this could be evidence that genomic loci associated with a complex trait across mammals may also be associated with variations in related traits within the human population.

The bulk motor cortex OCRs with lower predicted open chromatin in vocal learners show the strongest tendency to overlap with OCRs in Layer V ET neurons, which form long range projections (Fig. 4E). These results are consistent with our previous findings showing decreases in the expression of axon guidance genes in the motor cortex of vocal learning species (4). Among other functions, the Layer V ET neurons implicated by TACIT create the corticospinal projections that have been hypothesized as an anatomical landmark of vocal learners (2, 3, 48, 49, 56, 85, 86). Furthermore, the neuroethological and anatomical experiments we conducted in *R. Aegyptiacus* provide evidence that corticospinal projection neurons are present in the motor cortex of that bat species and that this motor cortical region participates in vocal production. Thus, consistent with previous literature, our results support a model in which the loss of regulatory element activity in the motor cortex influences axon guidance properties of long range projection neurons, which allow more robust connectivity between the cortex and the brainstem of vocal learning mammals (4, 85, 87). Alternatively, these genetic differences could relate to potential differences in the density of disynaptic connections that have been associated with skilled motor behavior, including vocalization in non-human primates (88, 89). Notably, these long range projection neurons have also been associated with predisposition to autism (90).

Materials and methods summary

To find vocal learning-associated convergent evolution in protein-coding sequences of the mammalian genome, we began with amino acid level multiple sequence alignments produced by the Zoonomia consortium (10). Those served as input to two classes

of methods, RERconverge (11) and HyPhy (30). RERconverge with an additional permutations correction for phylogenetic structure (24) was used to find protein-coding sequences whose evolutionary rates were associated with the presence or absence of vocal learning. HyPhy RELAX was used to find protein-coding sequences that were evolving more slowly, neutrally, or faster in vocal learning species. In addition, the HyPhy BUSTED-PH method (12) was applied to find evidence of diversifying positive selection. The gene ontology analysis was performed on the intersection of the RERconverge and HyPhy results using EnrichR (91). To control for false positives across all methods, Benjamini-Hochberg false discovery rate correction (92) was applied.

To examine the existence of a direct monosynaptic projection in a vocal learning mammal, the corticobulbar projections in Egyptian fruit-bats were mapped by tracing the projection from the orofacial motor cortex and from the cricothyroid muscles of the vocal cords. Performing immunohistochemistry in the brainstem revealed that synaptic boutons of cortical projection neurons overlapped with retrogradely-labeled motoneurons - confirming the existence of a direct monosynaptic projection. The role of the orofacial motor cortex during vocal production was then validated by quantifying the information between the vocalization amplitude and single cortical neuron activity measured wirelessly in vivo while the bats produced and listened to vocalizations.

To create an atlas of open chromatin regions in the bat motor cortex (M1), several brain regions, including wing-M1 and orofacial-M1, were separately dissected. The samples were cryopreserved, then the nuclei were isolated, and subsequently ATAC-Seq was performed to measure open chromatin. The open chromatin regions from this experiment were combined with previously published experiments in macaque, rat (23), and mouse (21) to create an atlas of cross-species motor cortex open chromatin.

To find vocal learning-associated convergent evolution in noncoding regions of the Boreoeutherian mammalian genome, the TACIT machine learning approach, was applied. Orthologous regions across genomes were found by combining the CACTUS whole genome multiple sequence alignment (19), hallLiftover (93) and HALPER (94). PhyloGlm (72) was then used to associate predicted motor cortex and parvalbumin-positive inhibitory interneuron open chromatin with binary annotations of vocal learning behavior. Phylogenetic permutations were applied to correct for phylogenetic tree structure and Benjamini-Hochberg to correct for multiple hypothesis testing. To identify potential trends in the cell type-specificity of the implicated regions, permutations on the regions of the genome that were predicted to have significantly higher or lower open chromatin in vocal learning species were conducted.

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). [doi:10.1146/annurev.neuro.22.1.567](https://doi.org/10.1146/annurev.neuro.22.1.567) [Medline](#)
2. M. Wirthlin, E. F. Chang, M. Knörnschild, L. A. Krubitzer, C. V. Mello, C. T. Miller,

- A. R. Pfenning, S. C. Vernes, O. Tchernichovski, M. M. Yartsev, A Modular Approach to Vocal Learning: Disentangling the Diversity of a Complex Behavioral Trait. *Neuron* **104**, 87–99 (2019). [doi:10.1016/j.neuron.2019.09.036](https://doi.org/10.1016/j.neuron.2019.09.036) [Medline](#)
3. E. D. Jarvis, Learned birdsong and the neurobiology of human language. *Ann. N. Y. Acad. Sci.* **1016**, 749–777 (2004). [doi:10.1196/annals.1298.038](https://doi.org/10.1196/annals.1298.038) [Medline](#)
 4. A. R. Pfenning, E. Hara, O. Whitney, M. V. Rivas, R. Wang, P. L. Roulhac, J. T. Howard, M. Wirthlin, P. V. Lovell, G. Ganapathy, J. Mouncastle, M. A. Moseley, J. W. Thompson, E. J. Soderblom, A. Iriki, M. Kato, M. T. P. Gilbert, G. Zhang, T. Bakken, A. Bongaarts, A. Bernard, E. Lein, C. V. Mello, A. J. Hartemink, E. D. Jarvis, Convergent transcriptional specializations in the brains of humans and song-learning birds. *Science* **346**, 1256846 (2014). [doi:10.1126/science.1256846](https://doi.org/10.1126/science.1256846) [Medline](#)
 5. G. Arriaga, E. D. Jarvis, Mouse vocal communication system: Are ultrasounds learned or innate? *Brain Lang.* **124**, 96–116 (2013). [doi:10.1016/j.bandl.2012.10.002](https://doi.org/10.1016/j.bandl.2012.10.002) [Medline](#)
 6. W. A. Searcy, J. Soha, S. Peters, S. Nowicki, Variation in vocal production learning across songbirds. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **376**, 20200257 (2021). [doi:10.1098/rstb.2020.0257](https://doi.org/10.1098/rstb.2020.0257) [Medline](#)
 7. P. L. Tyack, A taxonomy for vocal learning. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **375**, 20180406 (2020). [doi:10.1098/rstb.2018.0406](https://doi.org/10.1098/rstb.2018.0406) [Medline](#)
 8. H. Ruch, Y. Zürcher, J. M. Burkart, The function and mechanism of vocal accommodation in humans and other primates. *Biol. Rev. Camb. Philos. Soc.* **93**, 996–1013 (2018). [doi:10.1111/brv.12382](https://doi.org/10.1111/brv.12382) [Medline](#)
 9. Zoonomia Consortium, A comparative genomics multitool for scientific discovery and conservation. *Nature* **587**, 240–245 (2020). [doi:10.1038/s41586-020-2876-6](https://doi.org/10.1038/s41586-020-2876-6) [Medline](#)
 10. B. M. Kirilenko, C. Munegowda, E. Osipova, D. Jebb, V. Sharma, M. Blumer, A. E. Morales, A.-W. Ahmed, D.-G. Kontopoulos, L. Hilgers, K. Lindblad-Toh, E. K. Karlsson, M. Hiller; Zoonomia Consortium, Integrating gene annotation with orthology inference at scale. *Science* **380**, eabn3107 (2023). [doi:10.1126/science.abn3107](https://doi.org/10.1126/science.abn3107) [Medline](#)
 11. A. Kowalczyk, W. K. Meyer, R. Partha, W. Mao, N. L. Clark, M. Chikina, RERconverge: An R package for associating evolutionary rates with convergent traits. *Bioinformatics* **35**, 4815–4817 (2019). [doi:10.1093/bioinformatics/btz468](https://doi.org/10.1093/bioinformatics/btz468) [Medline](#)
 12. S. L. Kosakovsky Pond, A. F. Y. Poon, R. Velazquez, S. Weaver, N. L. Hepler, B. Murrell, S. D. Shank, B. R. Magalis, D. Bouvier, A. Nekrutenko, S. Wisotsky, S. J. Spielman, S. D. W. Frost, S. V. Muse, HyPhy 2.5-A Customizable Platform for Evolutionary Hypothesis Testing Using Phylogenies. *Mol. Biol. Evol.* **37**, 295–299 (2020). [doi:10.1093/molbev/msz197](https://doi.org/10.1093/molbev/msz197) [Medline](#)
 13. M. Tsompana, M. J. Buck, Chromatin accessibility: A window into the genome. *Epigenetics Chromatin* **7**, 33 (2014). [doi:10.1186/1756-8935-7-33](https://doi.org/10.1186/1756-8935-7-33) [Medline](#)
 14. Y. Prat, M. Taub, Y. Yovel, Vocal learning in a social mammal: Demonstrated by isolation and playback experiments in bats. *Sci. Adv.* **1**, e1500019 (2015). [doi:10.1126/sciadv.1500019](https://doi.org/10.1126/sciadv.1500019) [Medline](#)
 15. Y. Prat, L. Azoulay, R. Dor, Y. Yovel, Crowd vocal learning induces vocal dialects in bats: Playback of conspecifics shapes fundamental frequency usage by pups. *PLOS Biol.* **15**, e2002556 (2017). [doi:10.1371/journal.pbio.2002556](https://doi.org/10.1371/journal.pbio.2002556) [Medline](#)
 16. D. Genzel, J. Desai, E. Paras, M. M. Yartsev, Long-term and persistent vocal plasticity in adult bats. *Nat. Commun.* **10**, 3372 (2019). [doi:10.1038/s41467-019-11350-2](https://doi.org/10.1038/s41467-019-11350-2) [Medline](#)
 17. D. P. Genereux, A. Seres, J. Armstrong, J. Johnson, V. D. Marinescu, E. Murén, D. Juan, G. Bejerano, N. R. Casewell, L. G. Chemnick, F. Di Palma, M. Diekhans, I. T. Fiddes, V. N. Garber, V. N. Gladyshev, L. Goodman, W. Haerty, M. L. Houck, R. Hublely, T. Kivioja, K.-P. Koepfli, L. F. K. Kuderna, E. S. Lander, J. Meadows, W. J. Murphy, W. Nash, H. J. Noh, M. Nweeia, A. R. Pfenning, K. S. Pollard, D. Ray, B. Shapiro, A. Smit, M. Springer, C. C. Steiner, R. Swofford, J. Taipale, E. C. Teeling, J. Turner-Maier, J. Alföldi, B. Birren, O. A. Ryder, H. Lewin, B. Paten, T. Marques-Bonet, K. Lindblad-Toh, E. K. Karlsson; Zoonomia Consortium, A comparative genomics multitool for scientific discovery and conservation. *Nature* **587**, 240–245 (2020). [doi:10.1038/s41586-020-2876-6](https://doi.org/10.1038/s41586-020-2876-6) [Medline](#)
 18. M. J. Christmas, I. M. Kaplow, D. P. Genereux, M. X. Dong, G. M. Hughes, X. Li, P. F. Sullivan, A. G. Hindle, G. Andrews, J. C. Armstrong, M. Bianchi, A. M. Breit, M. Diekhans, C. Fanter, N. M. Foley, D. B. Goodman, L. Goodman, K. C. Keough, B. Kirilenko, A. Kowalczyk, C. Lawless, A. L. Lind, J. R. S. Meadows, L. R. Moreira, R. W. Redlich, L. Ryan, R. Swofford, A. Valenzuela, F. Wagner, O. Wallerman, A. R. Brown, J. Damas, K. Fan, J. Gatesy, J. Grimshaw, J. Johnson, S. V. Kozyrev, A. J. Lawler, V. D. Marinescu, K. M. Morrill, A. Osmanski, N. S. Paulat, B. N. Phan, S. K. Reilly, D. E. Schäffer, C. Steiner, M. A. Supple, A. P. Wilder, M. E. Wirthlin, J. R. Xue, B. W. Birren, S. Gazal, R. M. Hublely, K.-P. Koepfli, T. Marques-Bonet, W. K. Meyer, M. Nweeia, P. C. Sabeti, B. Shapiro, A. F. A. Smit, M. Springer, E. Teeling, Z. Weng, M. Hiller, D. L. Levesque, H. A. Lewin, W. J. Murphy, A. Navarro, B. Paten, K. S. Pollard, D. A. Ray, I. Ruf, O. A. Ryder, A. R. Pfenning, K. Lindblad-Toh, E. K. Karlsson; Zoonomia Consortium, Evolutionary constraint and innovation across hundreds of placental mammals. *Science* **380**, eabn3943 (2023). [doi:10.1123/science.abn3943](https://doi.org/10.1123/science.abn3943) [Medline](#)
 19. J. Armstrong, G. Hickey, M. Diekhans, I. T. Fiddes, Progressive Cactus is a multiple-genome aligner for the thousand-genome era. *Nature* **587**, 246–251 (2020).
 20. M. E. Wirthlin, I. M. Kaplow, A. J. Lawler, J. He, B. D. N. Phan, A. R. Brown, W. R. Stauffer, A. R. Pfenning, The Regulatory Evolution of the Primate Fine-Motor System, *bioRxiv* 2020.10.27.356733 [Preprint] (2020); <https://doi.org/10.1101/2020.10.27.356733>.
 21. C. Srinivasan, B. N. Phan, A. J. Lawler, E. Ramamurthy, M. Kleyman, A. R. Brown, I. M. Kaplow, M. E. Wirthlin, A. R. Pfenning, Addiction-Associated Genetic Variants Implicate Brain Cell Type- and Region-Specific Cis-Regulatory Elements in Addiction Neurobiology. *J. Neurosci.* **41**, 9008–9030 (2021). [doi:10.1523/JNEUROSCI.2534-20.2021](https://doi.org/10.1523/JNEUROSCI.2534-20.2021) [Medline](#)
 22. T. E. Bakken, N. L. Jorstad, Q. Hu, B. B. Lake, W. Tian, B. E. Kalmbach, M. Crow, R. D. Hodge, F. M. Krienen, S. A. Sorensen, J. Eggertmont, Z. Yao, B. D. Aevermann, A. I. Aldridge, A. Bartlett, D. Bertagnolli, T. Casper, R. G. Castanon, K. Crichton, T. L. Daigle, R. Dalley, N. Dee, N. Dembrow, D. Diep, S.-L. Ding, W. Dong, R. Fang, S. Fischer, M. Goldman, J. Goldy, L. T. Graybuck, B. R. Herb, X. Hou, J. Kancherla, M. Kroll, K. Lathia, B. van Lew, Y. E. Li, C. S. Liu, H. Liu, J. D. Lucero, A. Mahurkar, D. McMillen, J. A. Miller, M. Moussa, J. R. Nery, P. R. Nicovich, S.-Y. Niu, J. Orvis, J. K. Osteen, S. Owen, C. R. Palmer, T. Pham, N. Plongthongkum, O. Poirion, N. M. Reed, C. Rimorin, A. Rivkin, W. J. Romanow, A. E. Sedeño-Cortés, K. Siletti, S. Somasundaram, J. Sulc, M. Tieu, A. Torkelson, H. Tung, X. Wang, F. Xie, A. M. Yanny, R. Zhang, S. A. Ament, M. M. Behrens, H. C. Bravo, J. Chun, A. Dobin, J. Gillis, R. Hertzano, P. R. Hof, T. Höllt, G. D. Horwitz, C. D. Keene, P. V. Kharchenko, A. L. Ko, B. P. Lelieveldt, C. Luo, E. A. Mukamel, A. Pinto-Duarte, S. Preissl, A. Regev, B. Ren, R. H. Scheuermann, K. Smith, W. J. Spain, O. R. White, C. Koch, M. Hawrylycz, B. Tasic, E. Z. Macosko, S. A. McCarroll, J. T. Ting, H. Zeng, K. Zhang, G. Feng, J. R. Ecker, S. Linnarsson, E. S. Lein, Comparative cellular analysis of motor cortex in human, marmoset and mouse. *Nature* **598**, 111–119 (2021). [doi:10.1038/s41586-021-03465-8](https://doi.org/10.1038/s41586-021-03465-8) [Medline](#)
 23. I. M. Kaplow, A. J. Lawler, D. E. Schäffer, C. Srinivasan, H. H. Sestili, M. E. Wirthlin, B. N. Phan, K. Prasad, A. R. Brown, X. Zhang, K. Foley, D. P. Genereux, E. K. Karlsson, K. Lindblad-Toh, W. K. Meyer, A. R. Pfenning; Zoonomia Consortium, Relating enhancer genetic variation across mammals to complex phenotypes using machine learning. *Science* **380**, eabm7993 (2023). [doi:10.1126/science.abm7993](https://doi.org/10.1126/science.abm7993) [Medline](#)
 24. E. Saputra, A. Kowalczyk, L. Cusick, N. Clark, M. Chikina, Phylogenetic Permutations: A Statistically Rigorous Approach to Measure Confidence in Associations in a Phylogenetic Context. *Mol. Biol. Evol.* **38**, 3004–3021 (2021). [doi:10.1093/molbev/msab068](https://doi.org/10.1093/molbev/msab068) [Medline](#)
 25. 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). [doi:10.1002/cne.20201](https://doi.org/10.1002/cne.20201) [Medline](#)
 26. Z. Shi, Z. Piccus, X. Zhang, H. Yang, H. Jarrell, Y. Ding, Z. Teng, O. Tchernichovski, X. Li, miR-9 regulates basal ganglia-dependent developmental vocal learning and adult vocal performance in songbirds. *eLife* **7**, e29087 (2018). [doi:10.7554/eLife.29087](https://doi.org/10.7554/eLife.29087) [Medline](#)
 27. K. E. Savell, M. E. Zipperly, J. J. Tuscher, C. G. Duke, R. A. Phillips, A. J. Bauman, S. Thukral, F. A. Sultan, N. A. Goska, L. Ianov, J. J. Day, A dopamine-induced gene expression signature regulates neuronal function and cocaine response,

- doi:[10.1101/781872](https://doi.org/10.1101/781872).
28. J. He, M. Kleyman, J. Chen, A. Alikaya, K. M. Rothenhoefer, B. E. Ozturk, M. Wirthlin, A. C. Bostan, K. Fish, L. C. Byrne, A. R. Pfenning, W. R. Stauffer, Transcriptional and anatomical diversity of medium spiny neurons in the primate striatum. *Curr. Biol.* **31**, 5473–5486.e6 (2021). doi:[10.1016/j.cub.2021.10.015](https://doi.org/10.1016/j.cub.2021.10.015) [Medline](#)
 29. P. Bonnnavion, C. Varin, G. Fakhfour, P. M. Olondo, A. De Groot, A. Cornil, R. L. Lopez, E. P. Fernandez, E. Isingrini, Q. Rainer, K. Xu, E. Tzavara, E. Vigneault, S. Dumas, A. de Kerchove d'Exaerde, B. Giros, Unexpected contributions of striatal projection neurons coexpressing dopamine D1 and D2 receptors in balancing motor control. *bioRxiv* (2023), p. 2022.04.05.487163.
 30. J. O. Wertheim, B. Murrell, M. D. Smith, S. L. Kosakovsky Pond, K. Scheffler, RELAX: Detecting relaxed selection in a phylogenetic framework. *Mol. Biol. Evol.* **32**, 820–832 (2015). doi:[10.1093/molbev/msu400](https://doi.org/10.1093/molbev/msu400) [Medline](#)
 31. N. Unger, S. Heim, D. I. Hilger, S. Bludau, P. Pieperhoff, S. Cichon, K. Amunts, T. W. Mühlisen, Identification of Phonology-Related Genes and Functional Characterization of Broca's and Wernicke's Regions in Language and Learning Disorders. *Front. Neurosci.* **15**, 680762 (2021). doi:[10.3389/fnins.2021.680762](https://doi.org/10.3389/fnins.2021.680762) [Medline](#)
 32. D. J. Josifova, G. R. Monroe, F. Tessadori, E. de Graaff, B. van der Zwaag, S. G. Mehta, M. Harakalova, K. J. Duran, S. M. Savelberg, I. J. Nijman, H. Jungbluth, C. C. Hoogenraad, J. Bakkers, N. V. Knoers, H. V. Firth, P. L. Beales, G. van Haften, M. M. van Haelst, M. M. van Haelst; DDD Study, Heterozygous KIDINS220/ARMS nonsense variants cause spastic paraplegia, intellectual disability, nystagmus, and obesity. *Hum. Mol. Genet.* **25**, 2158–2167 (2016). doi:[10.1093/hmg/ddw082](https://doi.org/10.1093/hmg/ddw082) [Medline](#)
 33. K. Tossell, L. C. Andreea, C. Cudmore, E. Lang, U. Muthukrishnan, A. Lumsden, J. D. Gilthorpe, C. Irving, Lrrn1 is required for formation of the midbrain-hindbrain boundary and organiser through regulation of affinity differences between midbrain and hindbrain cells in chick. *Dev. Biol.* **352**, 341–352 (2011). doi:[10.1016/j.ydbio.2011.02.002](https://doi.org/10.1016/j.ydbio.2011.02.002) [Medline](#)
 34. C. Liu, Q. Hu, J. Jing, Y. Zhang, J. Jin, L. Zhang, L. Mu, Y. Liu, B. Sun, T. Zhang, Q. Kong, G. Wang, D. Wang, Y. Zhang, X. Liu, W. Zhao, J. Wang, T. Feng, H. Li, Regulator of G protein signaling 5 (RGS5) inhibits sonic hedgehog function in mouse cortical neurons. *Mol. Cell. Neurosci.* **83**, 65–73 (2017). doi:[10.1016/j.mcn.2017.06.005](https://doi.org/10.1016/j.mcn.2017.06.005) [Medline](#)
 35. Y.-Y. Han, J.-W. Zhou, Z.-W. Guo, Z.-Q. Wu, Z.-Y. Zhang, D.-X. Liu, C. Long, Multiple brain regions are involved in reaction to acute restraint stress in CYLD-knockout mice. *Stress* **26**, 2228925 (2023). doi:[10.1080/10253890.2023.2228925](https://doi.org/10.1080/10253890.2023.2228925) [Medline](#)
 36. P. Adak, N. Banerjee, S. Sinha, A. K. Bandyopadhyay, Gamma-Aminobutyric Acid Type A Receptor Variants are Associated with Autism Spectrum Disorders. *J. Mol. Neurosci.* **73**, 237–249 (2023). doi:[10.1007/s12031-023-02113-2](https://doi.org/10.1007/s12031-023-02113-2) [Medline](#)
 37. M. Mennesson, E. Rydgren, T. Lipina, E. Sokolowska, N. Kulenskaya, F. Morello, E. Ivakine, V. Voikar, V. Risbrough, J. Partanen, I. Hovatta, Kainate receptor auxiliary subunit NETO2 is required for normal fear expression and extinction. *Neuropsychopharmacology* **44**, 1855–1866 (2019). doi:[10.1038/s41386-019-0344-5](https://doi.org/10.1038/s41386-019-0344-5) [Medline](#)
 38. L. Gueneau, R. J. Fish, H. E. Shamseldin, N. Voisin, F. Tran Mau-Them, E. Preiksaitiene, G. R. Monroe, A. Lai, A. Putoux, F. Alias, Q. Ambusaidi, L. Ambrozaityte, L. Cimbalistienė, J. Delafontaine, N. Guex, M. Hashem, W. Kurdi, S. S. Jamuar, L. J. Ying, C. Bonnard, T. Pippucci, S. Pradervand, B. Roechert, P. M. van Hasselt, M. Wiederkehr, C. F. Wright, I. Xenarios, G. van Haften, C. Shaw-Smith, E. M. Schindewolf, M. Neerman-Arbez, D. Sanlaville, G. Lesca, L. Guibaud, B. Reversade, J. Chelly, V. Kučinskas, F. S. Alkuraya, A. Raymond; DDD Study, KIAA1109 Variants Are Associated with a Severe Disorder of Brain Development and Arthrogyrosis. *Am. J. Hum. Genet.* **102**, 116–132 (2018). doi:[10.1016/j.ajhg.2017.12.002](https://doi.org/10.1016/j.ajhg.2017.12.002) [Medline](#)
 39. A. Gialluisi, D. F. Newbury, E. G. Willcutt, R. K. Olson, J. C. DeFries, W. M. Brandler, B. F. Pennington, S. D. Smith, T. S. Scerri, N. H. Simpson, M. Luciano, D. M. Evans, T. C. Bates, J. F. Stein, J. B. Talcott, A. P. Monaco, S. Paracchini, C. Francks, S. E. Fisher, S. E. Fisher; SLI Consortium, Genome-wide screening for DNA variants associated with reading and language traits. *Genes Brain Behav.* **13**, 686–701 (2014). doi:[10.1111/gbb.12158](https://doi.org/10.1111/gbb.12158) [Medline](#)
 40. A. K. Adams, S. D. Smith, D. T. Truong, E. G. Willcutt, R. K. Olson, J. C. DeFries, B. F. Pennington, J. R. Gruen, Enrichment of putatively damaging rare variants in the DYX2 locus and the reading-related genes CCDC136 and FLNC. *Hum. Genet.* **136**, 1395–1405 (2017). doi:[10.1007/s00439-017-1838-z](https://doi.org/10.1007/s00439-017-1838-z) [Medline](#)
 41. H. Tager-Flusberg, R. Paul, C. Lord, “Language and communication in autism” in *Handbook of autism and pervasive developmental disorders*, vol. 1, F. R. Volkmar, R. Paul, A. Klin, D. Cohen, Eds. (Wiley, ed. 3, 2005), pp. 335–364.
 42. E. Schoen, R. Paul, K. Chawarska, Phonology and vocal behavior in toddlers with autism spectrum disorders. *Autism Res.* **4**, 177–188 (2011). doi:[10.1002/aur.183](https://doi.org/10.1002/aur.183) [Medline](#)
 43. P. M. Rodenas-Cuadrado, J. Mengede, L. Baas, P. Devanna, T. A. Schmid, M. Yartsev, U. Firzlaff, S. C. Vernes, Mapping the distribution of language related genes FoxP1, FoxP2, and Cntna2 in the brains of vocal learning bat species. *J. Comp. Neurol.* **526**, 1235–1266 (2018). doi:[10.1002/cne.24385](https://doi.org/10.1002/cne.24385) [Medline](#)
 44. J. A. Cahill, J. Armstrong, A. Deran, C. J. Khoury, B. Paten, D. Haussler, E. D. Jarvis, Positive selection in noncoding genomic regions of vocal learning birds is associated with genes implicated in vocal learning and speech functions in humans. *Genome Res.* **31**, 2035–2049 (2021). doi:[10.1101/gr.275989.121](https://doi.org/10.1101/gr.275989.121) [Medline](#)
 45. B. St Pourcain, R. A. M. Cents, A. J. O. Whitehouse, C. M. A. Haworth, O. S. P. Davis, P. F. O'Reilly, S. Roulstone, Y. Wren, Q. W. Ang, F. P. Velders, D. M. Evans, J. P. Kemp, N. M. Warrington, L. Miller, N. J. Timpson, S. M. Ring, F. C. Verhulst, A. Hofman, F. Rivadeneira, E. L. Meaburn, T. S. Price, P. S. Dale, D. Pillas, A. Ylihera, A. Rodriguez, J. Golding, V. W. V. Jaddoe, M.-R. Jarvelin, R. Plomin, C. E. Pennell, H. Tiemeier, G. Davey Smith, Common variation near ROBO2 is associated with expressive vocabulary in infancy. *Nat. Commun.* **5**, 4831 (2014). doi:[10.1038/ncomms5831](https://doi.org/10.1038/ncomms5831) [Medline](#)
 46. M. C. Rose, B. Styr, T. A. Schmid, J. E. Elie, M. M. Yartsev, Cortical representation of group social communication in bats. *Science* **374**, eaba9584 (2021). doi:[10.1126/science.aba9584](https://doi.org/10.1126/science.aba9584) [Medline](#)
 47. A. C. Halley, M. K. L. Baldwin, D. F. Cooke, M. Englund, C. R. Pineda, T. Schmid, M. M. Yartsev, L. Krubitzer, Coevolution of motor cortex and behavioral specializations associated with flight and echolocation in bats. *Curr. Biol.* **32**, 2935–2941.e3 (2022). doi:[10.1016/j.cub.2022.04.094](https://doi.org/10.1016/j.cub.2022.04.094) [Medline](#)
 48. K. Simonyan, The laryngeal motor cortex: Its organization and connectivity. *Curr. Opin. Neurobiol.* **28**, 15–21 (2014). doi:[10.1016/j.conb.2014.05.006](https://doi.org/10.1016/j.conb.2014.05.006) [Medline](#)
 49. U. Jürgens, Neural pathways underlying vocal control. *Neurosci. Biobehav. Rev.* **26**, 235–258 (2002). doi:[10.1016/S0149-7634\(01\)00068-9](https://doi.org/10.1016/S0149-7634(01)00068-9) [Medline](#)
 50. K. Simonyan, B. Horwitz, Laryngeal motor cortex and control of speech in humans. *Neuroscientist* **17**, 197–208 (2011). doi:[10.1177/1073858410386727](https://doi.org/10.1177/1073858410386727) [Medline](#)
 51. P. T. Martins, C. Boeckx, Vocal learning: Beyond the continuum. *PLOS Biol.* **18**, e3000672 (2020). doi:[10.1371/journal.pbio.3000672](https://doi.org/10.1371/journal.pbio.3000672) [Medline](#)
 52. W. T. Fitch, The Biology and Evolution of Speech: A Comparative Analysis. *Annu. Rev. Linguist.* **4**, 255–279 (2018). doi:[10.1146/annurev-linguistics-011817-045748](https://doi.org/10.1146/annurev-linguistics-011817-045748)
 53. J. R. Belkhir, W. T. Fitch, F. E. Garcea, B. L. Chernoff, M. H. Sims, E. Navarrete, S. Haber, D. A. Paul, S. O. Smith, W. H. Pilcher, B. Z. Mahon, Direct electrical stimulation evidence for a dorsal motor area with control of the larynx. *Brain Stimul.* **14**, 110–112 (2021). doi:[10.1016/j.brs.2020.11.013](https://doi.org/10.1016/j.brs.2020.11.013) [Medline](#)
 54. J. D. Breshears, A. M. Molinaro, E. F. Chang, A probabilistic map of the human ventral sensorimotor cortex using electrical stimulation. *J. Neurosurg.* **123**, 340–349 (2015). doi:[10.3171/2014.11.JNS14889](https://doi.org/10.3171/2014.11.JNS14889) [Medline](#)
 55. B. K. Dichter, J. D. Breshears, M. K. Leonard, E. F. Chang, The Control of Vocal Pitch in Human Laryngeal Motor Cortex. *Cell* **174**, 21–31.e9 (2018). doi:[10.1016/j.cell.2018.05.016](https://doi.org/10.1016/j.cell.2018.05.016) [Medline](#)
 56. H. G. J. M. Kuypers, Corticobulbar connexions to the pons and lower brain-stem in man: An anatomical study. *Brain* **81**, 364–388 (1958). doi:[10.1093/brain/81.3.364](https://doi.org/10.1093/brain/81.3.364) [Medline](#)
 57. J. M. Wild, Neural pathways for the control of birdsong production. *J. Neurobiol.* **33**, 653–670 (1997). doi:[10.1002/\(SICI\)1097-4695\(19971105\)33:5<653::AID-NEU11>3.0.CO;2-A](https://doi.org/10.1002/(SICI)1097-4695(19971105)33:5<653::AID-NEU11>3.0.CO;2-A) [Medline](#)

58. J. A. Paton, K. R. Manogue, F. Nottebohm, Bilateral organization of the vocal control pathway in the budgerigar, *Melopsittacus undulatus*. *J. Neurosci.* **1**, 1279–1288 (1981). [doi:10.1523/JNEUROSCI.01-11-01279.1981](https://doi.org/10.1523/JNEUROSCI.01-11-01279.1981) [Medline](#)
59. M. Gahr, Neural song control system of hummingbirds: Comparison to swifts, vocal learning (Songbirds) and nonlearning (Suboscines) passerines, and vocal learning (Budgerigars) and nonlearning (Dove, owl, gull, quail, chicken) nonpasserines. *J. Comp. Neurol.* **426**, 182–196 (2000). [doi:10.1002/1096-9861\(20001016\)426:2<182::AID-CNE2>3.0.CO;2-M](https://doi.org/10.1002/1096-9861(20001016)426:2<182::AID-CNE2>3.0.CO;2-M) [Medline](#)
60. D.-J. Zheng, D. E. Okobi Jr., R. Shu, R. Agrawal, S. K. Smith, M. A. Long, S. M. Phelps, Mapping the vocal circuitry of Alston’s singing mouse with pseudorabies virus. *J. Comp. Neurol.* **530**, 2075–2099 (2022). [doi:10.1002/cne.25321](https://doi.org/10.1002/cne.25321) [Medline](#)
61. J. D. Buenrostro, P. G. Giresi, L. C. Zaba, H. Y. Chang, W. J. Greenleaf, Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* **10**, 1213–1218 (2013). [doi:10.1038/nmeth.2688](https://doi.org/10.1038/nmeth.2688) [Medline](#)
62. S. E. Fisher, C. Scharff, FOXP2 as a molecular window into speech and language. *Trends Genet.* **25**, 166–177 (2009). [doi:10.1016/j.tig.2009.03.002](https://doi.org/10.1016/j.tig.2009.03.002) [Medline](#)
63. M. C. King, A. C. Wilson, Evolution at two levels in humans and chimpanzees. *Science* **188**, 107–116 (1975). [doi:10.1126/science.1090005](https://doi.org/10.1126/science.1090005) [Medline](#)
64. E. Z. Kvon, O. K. Kameeva, U. S. Melo, I. Barozzi, M. Osterwalder, B. J. Mannion, V. Tissières, C. S. Pickle, I. Plajzer-Frick, E. A. Lee, M. Kato, T. H. Garvin, J. A. Akiyama, V. Afzal, J. Lopez-Rios, E. M. Rubin, D. E. Dickel, L. A. Pennacchio, A. Visel, Progressive Loss of Function in a Limb Enhancer during Snake Evolution. *Cell* **167**, 633–642.e11 (2016). [doi:10.1016/j.cell.2016.09.028](https://doi.org/10.1016/j.cell.2016.09.028) [Medline](#)
65. R. Partha, B. K. Chauhan, Z. Ferreira, J. D. Robinson, K. Lathrop, K. K. Nischal, M. Chikina, N. L. Clark, Subterranean mammals show convergent regression in ocular genes and enhancers, along with adaptation to tunneling. *eLife* **6**, e25884 (2017). [doi:10.7554/eLife.25884](https://doi.org/10.7554/eLife.25884) [Medline](#)
66. E. S. Wong, D. Zheng, S. Z. Tan, N. L. Bower, V. Garside, G. Vanwalleghem, F. Gaiti, E. Scott, B. M. Hogan, K. Kikuchi, E. McGlenn, M. Francois, B. M. Degnan, Deep conservation of the enhancer regulatory code in animals. *Science* **370**, eaax8137 (2020). [doi:10.1126/science.aax8137](https://doi.org/10.1126/science.aax8137) [Medline](#)
67. I. M. Kaplow, D. E. Schäffer, M. E. Wirthlin, A. J. Lawler, A. R. Brown, M. Kleyman, A. R. Pfenning, Inferring mammalian tissue-specific regulatory conservation by predicting tissue-specific differences in open chromatin. *BMC Genomics* **23**, 291 (2022). [doi:10.1186/s12864-022-08450-7](https://doi.org/10.1186/s12864-022-08450-7) [Medline](#)
68. Roadmap Epigenomics Consortium, Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–330 (2015).
69. M. Hiller, B. T. Schaar, V. B. Indjeian, D. M. Kingsley, L. R. Hagey, G. Bejerano, A “forward genomics” approach links genotype to phenotype using independent phenotypic losses among related species. *Cell Rep.* **2**, 817–823 (2012). [doi:10.1016/j.celrep.2012.08.032](https://doi.org/10.1016/j.celrep.2012.08.032) [Medline](#)
70. K. S. Pollard, M. J. Hubisz, K. R. Rosenbloom, A. Siepel, Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Res.* **20**, 110–121 (2010). [doi:10.1101/gr.097857.109](https://doi.org/10.1101/gr.097857.109) [Medline](#)
71. Y. E. Li, S. Preissl, X. Hou, Z. Zhang, K. Zhang, Y. Qiu, O. B. Poirion, B. Li, J. Chiou, H. Liu, A. Pinto-Duarte, N. Kubo, X. Yang, R. Fang, X. Wang, J. Y. Han, J. Lucero, Y. Yan, M. Miller, S. Kuan, D. Gorkin, K. J. Gaulton, Y. Shen, M. Nunn, E. A. Mukamel, M. M. Behrens, J. R. Ecker, B. Ren, An atlas of gene regulatory elements in adult mouse cerebrum. *Nature* **598**, 129–136 (2021). [doi:10.1038/s41586-021-03604-1](https://doi.org/10.1038/s41586-021-03604-1) [Medline](#)
72. L. S. Tung Ho, C. Ané, A linear-time algorithm for Gaussian and non-Gaussian trait evolution models. *Syst. Biol.* **63**, 397–408 (2014). [doi:10.1093/sysbio/syu005](https://doi.org/10.1093/sysbio/syu005) [Medline](#)
73. A. R. Ives, T. Garland Jr., Phylogenetic logistic regression for binary dependent variables. *Syst. Biol.* **59**, 9–26 (2010). [doi:10.1093/sysbio/syp074](https://doi.org/10.1093/sysbio/syp074) [Medline](#)
74. BRAIN Initiative Cell Census Network (BICCN), A multimodal cell census and atlas of the mammalian primary motor cortex. *Nature* **598**, 86–102 (2021). [doi:10.1038/s41586-021-03950-0](https://doi.org/10.1038/s41586-021-03950-0) [Medline](#)
75. M. Munz, A. Bharioke, G. Kosche, V. Moreno-Juan, A. Brignall, T. M. Rodrigues, A. Graff-Meyer, T. Ulmer, S. Haeuselmann, D. Pavlinic, N. Ledergerber, B. Gross-Scherf, B. Rózsa, J. Krol, S. Picelli, C. S. Cowan, B. Roska, Pyramidal neurons form active, transient, multilayered circuits perturbed by autism-associated mutations at the inception of neocortex. *Cell* **186**, 1930–1949.e31 (2023). [doi:10.1016/j.cell.2023.03.025](https://doi.org/10.1016/j.cell.2023.03.025) [Medline](#)
76. R. Barone, M. Fichera, M. De Grandi, M. Battaglia, V. Lo Faro, T. Mattina, R. Rizzo, Familial 18q12.2 deletion supports the role of RNA-binding protein CELF4 in autism spectrum disorders. *Am. J. Med. Genet. A.* **173**, 1649–1655 (2017). [doi:10.1002/ajmg.a.38205](https://doi.org/10.1002/ajmg.a.38205) [Medline](#)
77. W. Sun, J. L. Wagnon, C. L. Mahaffey, M. Briese, J. Ule, W. N. Frankel, Aberrant sodium channel activity in the complex seizure disorder of Celf4 mutant mice. *J. Physiol.* **591**, 241–255 (2013). [doi:10.1113/jphysiol.2012.240168](https://doi.org/10.1113/jphysiol.2012.240168) [Medline](#)
78. E. K. Ruzzo, L. Pérez-Cano, J.-Y. Jung, L.-K. Wang, D. Kashef-Haghighi, C. Hartl, C. Singh, J. Xu, J. N. Hoekstra, O. Leventhal, V. M. Leppä, M. J. Gandal, K. Paskov, N. Stockham, D. Polioudakis, J. K. Lowe, D. A. Prober, D. H. Geschwind, D. P. Wall, Inherited and De Novo Genetic Risk for Autism Impacts Shared Networks. *Cell* **178**, 850–866.e26 (2019). [doi:10.1016/j.cell.2019.07.015](https://doi.org/10.1016/j.cell.2019.07.015) [Medline](#)
79. M. Niarchou, D. E. Gustavson, J. F. Sathirapongsasuti, M. Anglada-Tort, E. Eising, E. Bell, E. McArthur, P. Straub, J. D. McAuley, J. A. Capra, F. Ullén, N. Creanza, M. A. Mosing, D. A. Hinds, L. K. Davis, N. Jacoby, R. L. Gordon, P. M. Gandhi, K. Heilbron, B. Hicks, K. E. Huber, E. M. Jewett, Y. Jiang, A. Kleinman, K.-H. Lin, N. K. Litterman, J. C. McCreight, M. H. McIntyre, K. F. McManus, J. L. Mountain, S. V. Mozaffari, P. Nandakumar, E. S. Noblin, C. A. M. Northover, J. O’Connell, S. J. Pitts, G. D. Poznik, A. J. Shastri, J. F. Shelton, S. Shringarpure, C. Tian, J. Y. Tung, R. J. Tunney, V. Vacic, X. Wang, J. D. McAuley, J. A. Capra, F. Ullén, N. Creanza, M. A. Mosing, D. A. Hinds, L. K. Davis, N. Jacoby, R. L. Gordon; 23andMe Research Team, Genome-wide association study of musical beat synchronization demonstrates high polygenicity. *Nat. Hum. Behav.* **6**, 1292–1309 (2022). [doi:10.1038/s41562-022-01359-x](https://doi.org/10.1038/s41562-022-01359-x) [Medline](#)
80. T. Karayannis, E. Au, J. C. Patel, I. Kruglikov, S. Markx, R. Delorme, D. Héron, D. Salomon, J. Glessner, S. Restituito, A. Gordon, L. Rodriguez-Murillo, N. C. Roy, J. A. Gogos, B. Rudy, M. E. Rice, M. Karayiorgou, H. Hakonarson, B. Keren, G. Huguet, T. Bourgeron, C. Hoeffler, R. W. Tsien, E. Peles, G. Fishell, Cntnap4 differentially contributes to GABAergic and dopaminergic synaptic transmission. *Nature* **511**, 236–240 (2014). [doi:10.1038/nature13248](https://doi.org/10.1038/nature13248) [Medline](#)
81. C. I. S. Costa, E. M. da Silva Montenegro, M. Zarrei, E. de Sá Moreira, I. M. W. Silva, M. de Oliveira Scliar, J. Y. T. Wang, E. C. Zachi, E. V. Branco, S. S. da Costa, N. C. V. Lourenço, A. M. Vianna-Morgante, C. Rosenberg, A. C. V. Krepisch, S. W. Scherer, M. R. Passos-Bueno, Copy number variations in a Brazilian cohort with autism spectrum disorders highlight the contribution of cell adhesion genes. *Clin. Genet.* **101**, 134–141 (2022). [doi:10.1111/cge.14072](https://doi.org/10.1111/cge.14072) [Medline](#)
82. B. S. Abrahams, D. E. Arking, D. B. Campbell, H. C. Mefford, E. M. Morrow, L. A. Weiss, I. Menashe, T. Wadkins, S. Banerjee-Basu, A. Packer, SFARI Gene 2.0: A community-driven knowledgebase for the autism spectrum disorders (ASDs). *Mol. Autism* **4**, 36 (2013). [doi:10.1186/2040-2392-4-36](https://doi.org/10.1186/2040-2392-4-36) [Medline](#)
83. P. F. Sullivan, J. R. S. Meadows, S. Gazal, B. N. Phan, X. Li, D. P. Genereux, M. X. Dong, M. Bianchi, G. Andrews, S. Sakhthikumar, J. Nordin, A. Roy, M. J. Christmas, V. D. Marinescu, C. Wang, O. Wallerman, J. Xue, S. Yao, Q. Sun, J. Szatkiewicz, J. Wen, L. M. Huckins, A. Lawler, K. C. Keough, Z. Zheng, J. Zeng, N. R. Wray, Y. Li, J. Johnson, J. Chen, B. Paten, S. K. Reilly, G. M. Hughes, Z. Weng, K. S. Pollard, A. R. Pfenning, K. Forsberg-Nilsson, E. K. Karlsson, K. Lindblad-Toh; Zoonomia Consortium, Leveraging base pair mammalian constraint to understand genetic variation and human disease. *Science* **380**, eabn2937 (2023). [doi:10.1126/science.abn2937](https://doi.org/10.1126/science.abn2937) [Medline](#)
84. A. Brignell, A. T. Morgan, S. Woolfenden, F. Klopper, T. May, V. Sarkozy, K. Williams, A systematic review and meta-analysis of the prognosis of language outcomes for individuals with autism spectrum disorder. *Autism Dev. Lang. Impair.* **3**, 2396941518767610 (2018). [doi:10.1177/2396941518767610](https://doi.org/10.1177/2396941518767610) [Medline](#)
85. A. A. Nevue, P. V. Lovell, M. Wirthlin, C. V. Mello, Molecular specializations of deep cortical layer analogs in songbirds. *Sci. Rep.* **10**, 18767 (2020). [doi:10.1038/s41598-020-75773-4](https://doi.org/10.1038/s41598-020-75773-4) [Medline](#)
86. E. D. Jarvis, Evolution of vocal learning and spoken language. *Science* **366**, 50–54 (2019). [doi:10.1126/science.aax0287](https://doi.org/10.1126/science.aax0287) [Medline](#)
87. R. Wang, C.-C. Chen, E. Hara, M. V. Rivas, P. L. Roulhac, J. T. Howard, M. Chakraborty, J.-N. Audet, E. D. Jarvis, Convergent differential regulation of SLIT-ROBO axon guidance genes in the brains of vocal learners. *J. Comp.*

- Neurosci.* **523**, 892–906 (2015). [doi:10.1002/cne.23719](https://doi.org/10.1002/cne.23719) [Medline](#)
88. P. L. Strick, R. P. Dum, J.-A. Rathelot, The Cortical Motor Areas and the Emergence of Motor Skills: A Neuroanatomical Perspective. *Annu. Rev. Neurosci.* **44**, 425–447 (2021). [doi:10.1146/annurev-neuro-070918-050216](https://doi.org/10.1146/annurev-neuro-070918-050216) [Medline](#)
 89. C. M. Cerkevich, J.-A. Rathelot, P. L. Strick, Cortical basis for skilled vocalization. *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2122345119 (2022). [doi:10.1073/pnas.2122345119](https://doi.org/10.1073/pnas.2122345119) [Medline](#)
 90. A. J. Willsey, S. J. Sanders, M. Li, S. Dong, A. T. Tebbenkamp, R. A. Muhle, S. K. Reilly, L. Lin, S. Fertuzinhos, J. A. Miller, M. T. Murtha, C. Bichsel, W. Niu, J. Cotney, A. G. Ercan-Sencicek, J. Gockley, A. R. Gupta, W. Han, X. He, E. J. Hoffman, L. Klei, J. Lei, W. Liu, L. Liu, C. Lu, X. Xu, Y. Zhu, S. M. Mane, E. S. Lein, L. Wei, J. P. Noonan, K. Roeder, B. Devlin, N. Sestan, M. W. State, Coexpression networks implicate human midfetal deep cortical projection neurons in the pathogenesis of autism. *Cell* **155**, 997–1007 (2013). [doi:10.1016/j.cell.2013.10.020](https://doi.org/10.1016/j.cell.2013.10.020) [Medline](#)
 91. Z. Xie, A. Bailey, M. V. Kuleshov, D. J. B. Clarke, J. E. Evangelista, S. L. Jenkins, A. Lachmann, M. L. Wojciechowicz, E. Kropiwnicki, K. M. Jagodnik, M. Jeon, A. Ma'ayan, Gene Set Knowledge Discovery with Enrichr. *Curr. Protoc.* **1**, e90 (2021). [doi:10.1002/cpz1.90](https://doi.org/10.1002/cpz1.90) [Medline](#)
 92. Y. Benjamini, Y. Hochberg, Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc.* **57**, 289–300 (1995). [doi:10.1111/j.2517-6161.1995.tb02031.x](https://doi.org/10.1111/j.2517-6161.1995.tb02031.x)
 93. G. Hickey, B. Paten, D. Earl, D. Zerbin, D. Haussler, HAL: A hierarchical format for storing and analyzing multiple genome alignments. *Bioinformatics* **29**, 1341–1342 (2013). [doi:10.1093/bioinformatics/btt128](https://doi.org/10.1093/bioinformatics/btt128) [Medline](#)
 94. X. Zhang, I. M. Kaplow, M. Wirthlin, T. Y. Park, A. R. Pfening, HALPER facilitates the identification of regulatory element orthologs across species. *Bioinformatics* **36**, 4339–4340 (2020). [doi:10.1093/bioinformatics/btaa493](https://doi.org/10.1093/bioinformatics/btaa493) [Medline](#)
 95. R. Redlich, A. Kowalczyk, A. R. Pfening, pfenninglab/RErconvergeVocalLearning: Vocal Learning Mammalian Protein, Zenodo (2024); <https://zenodo.org/doi/10.5281/zenodo.10641176>.
 96. I. M. Kaplow, D. Schaeffer, C. Srinivasan, H. H. Sestilli, A. J. Lawler, A. R. Pfening, pfenninglab/TACIT: TACIT Vocal Learning Mammal M1+PV, Zenodo (2024); <https://zenodo.org/doi/10.5281/zenodo.5952292>.
 97. N. M. Foley, V. C. Mason, A. J. Harris, K. R. Bredemeyer, J. Damas, H. A. Lewin, E. Eizirik, J. Gates, M. S. Springer, W. J. Murphy; Zoonomia Consortium, A genomic timescale for placental mammal evolution. *Science* **380**, eabl8189 (2023).
 98. J. Armstrong, G. Hickey, M. Diekhans, I. T. Fiddes, A. M. Novak, A. Deran, Q. Fang, D. Xie, S. Feng, J. Stillier, D. Genereux, J. Johnson, V. D. Marinescu, J. Alföldi, R. S. Harris, K. Lindblad-Toh, D. Haussler, E. Karlsson, E. D. Jarvis, G. Zhang, B. Paten, Progressive Cactus is a multiple-genome aligner for the thousand-genome era. *Nature* **587**, 246–251 (2020). [doi:10.1038/s41586-020-2871-y](https://doi.org/10.1038/s41586-020-2871-y) [Medline](#)
 99. V. M. Janik, M. Knörnschild, Vocal production learning in mammals revisited. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **376**, 20200244 (2021). [doi:10.1098/rstb.2020.0244](https://doi.org/10.1098/rstb.2020.0244) [Medline](#)
 100. T. O'Rourke, P. T. Martins, R. Asano, R. O. Tachibana, K. Okanoya, C. Boeckx, Capturing the Effects of Domestication on Vocal Learning Complexity. *Trends Cogn. Sci.* **25**, 462–474 (2021). [doi:10.1016/j.tics.2021.03.007](https://doi.org/10.1016/j.tics.2021.03.007) [Medline](#)
 101. M. Chikina, J. D. Robinson, N. L. Clark, Hundreds of Genes Experienced Convergent Shifts in Selective Pressure in Marine Mammals. *Mol. Biol. Evol.* **33**, 2182–2192 (2016). [doi:10.1093/molbev/msw112](https://doi.org/10.1093/molbev/msw112) [Medline](#)
 102. B. M. Kirilenko, C. Munegowda, E. Osipova, D. Jebb, V. Sharma, M. Blumer, A. E. Morales, A.-W. Ahmed, D.-G. Kontopoulos, L. Hilgers, K. Lindblad-Toh, E. K. Karlsson, M. Hiller; Zoonomia Consortium, Integrating gene annotation with orthology inference at scale. *Science* **380**, eabn3107 (2023). [doi:10.1126/science.abn3107](https://doi.org/10.1126/science.abn3107)
 103. S. Whelan, N. Goldman, A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Mol. Biol. Evol.* **18**, 691–699 (2001). [doi:10.1093/oxfordjournals.molbev.a003851](https://doi.org/10.1093/oxfordjournals.molbev.a003851) [Medline](#)
 104. K. P. Schliep, phangorn: Phylogenetic analysis in R. *Bioinformatics* **27**, 592–593 (2011). [doi:10.1093/bioinformatics/btq706](https://doi.org/10.1093/bioinformatics/btq706) [Medline](#)
 105. Y. Benjamini, Y. Hochberg, Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. R. Stat. Soc.* **57**, 289–300 (1995).
 106. S. R. Wisotsky, S. L. Kosakovsky Pond, S. D. Shank, S. V. Muse, Synonymous Site-to-Site Substitution Rate Variation Dramatically Inflates False Positive Rates of Selection Analyses: Ignore at Your Own Peril. *Mol. Biol. Evol.* **37**, 2430–2439 (2020). [doi:10.1093/molbev/msaa037](https://doi.org/10.1093/molbev/msaa037) [Medline](#)
 107. M. M. Yartsev, M. P. Witter, N. Ulanovsky, Grid cells without theta oscillations in the entorhinal cortex of bats. *Nature* **479**, 103–107 (2011). [doi:10.1038/nature10583](https://doi.org/10.1038/nature10583) [Medline](#)
 108. D. Genzel, M. M. Yartsev, The fully automated bat (FAB) flight room: A human-free environment for studying navigation in flying bats and its initial application to the retrosplenial cortex. *J. Neurosci. Methods* **348**, 108970 (2021). [doi:10.1016/j.jneumeth.2020.108970](https://doi.org/10.1016/j.jneumeth.2020.108970) [Medline](#)
 109. W. L. Conte, H. Kamishina, R. L. Reep, Multiple neuroanatomical tract-tracing using fluorescent Alexa Fluor conjugates of cholera toxin subunit B in rats. *Nat. Protoc.* **4**, 1157–1166 (2009). [doi:10.1038/nprot.2009.93](https://doi.org/10.1038/nprot.2009.93) [Medline](#)
 110. S. Q. Lima, T. Hromádka, P. Znamenskiy, A. M. Zador, PINP: A new method of tagging neuronal populations for identification during in vivo electrophysiological recording. *PLOS ONE* **4**, e6099 (2009). [doi:10.1371/journal.pone.0006099](https://doi.org/10.1371/journal.pone.0006099) [Medline](#)
 111. D. Knowland, V. Lilascharoen, C. P. Pacia, S. Shin, E. H.-J. Wang, B. K. Lim, Distinct Ventral Pallidal Neural Populations Mediate Separate Symptoms of Depression. *Cell* **170**, 284–297.e18 (2017). [doi:10.1016/j.cell.2017.06.015](https://doi.org/10.1016/j.cell.2017.06.015) [Medline](#)
 112. Q. Cui, X. Du, I. Y. M. Chang, A. Pamukcu, V. Lilascharoen, B. L. Berceau, D. García, D. Hong, U. Chon, A. Narayanan, Y. Kim, B. K. Lim, C. S. Chan, Striatal Direct Pathway Targets Npas1⁺ Pallidal Neurons. *J. Neurosci.* **41**, 3966–3987 (2021). [doi:10.1523/JNEUROSCI.2306-20.2021](https://doi.org/10.1523/JNEUROSCI.2306-20.2021) [Medline](#)
 113. W. Zhang, M. M. Yartsev, Correlated Neural Activity across the Brains of Socially Interacting Bats. *Cell* **178**, 413–428.e22 (2019). [doi:10.1016/j.cell.2019.05.023](https://doi.org/10.1016/j.cell.2019.05.023) [Medline](#)
 114. N. A. Shneider, G. Z. Mentis, J. Schustak, M. J. O'Donovan, Functionally reduced sensorimotor connections form with normal specificity despite abnormal muscle spindle development: The role of spindle-derived neurotrophin 3. *J. Neurosci.* **29**, 4719–4735 (2009). [doi:10.1523/JNEUROSCI.5790-08.2009](https://doi.org/10.1523/JNEUROSCI.5790-08.2009) [Medline](#)
 115. J. E. Elie, F. E. Theunissen, The vocal repertoire of the domesticated zebra finch: A data-driven approach to decipher the information-bearing acoustic features of communication signals. *Anim. Cogn.* **19**, 285–315 (2016). [doi:10.1007/s10071-015-0933-6](https://doi.org/10.1007/s10071-015-0933-6) [Medline](#)
 116. C. Stringer, M. Pachitariu, N. Steinmetz, C. B. Reddy, M. Carandini, K. D. Harris, Spontaneous behaviors drive multidimensional, brainwide activity. *Science* **364**, 255 (2019). [doi:10.1126/science.aav7893](https://doi.org/10.1126/science.aav7893) [Medline](#)
 117. M. Pachitariu, N. A. Steinmetz, S. N. Kadir, M. Carandini, K. D. Harris, “Fast and accurate spike sorting of high-channel count probes with KiloSort” in *Advances in Neural Information Processing Systems* 29, D. Lee, M. Sugiyama, U. Luxburg, I. Guyon, R. Garnett, Eds. (NIPS, 2016); <https://proceedings.neurips.cc/paper/2016/hash/1145a30ff80745b56fb0cec6f5305017-Abstract.html>.
 118. MouseLand/Kilosort, Github; <https://github.com/MouseLand/Kilosort>.
 119. A. Hsu, A. Borst, F. E. Theunissen, Quantifying variability in neural responses and its application for the validation of model predictions. *Network* **15**, 91–109 (2004). [doi:10.1088/0954-898X/15/2/002](https://doi.org/10.1088/0954-898X/15/2/002) [Medline](#)
 120. J. D. Buenrostro, B. Wu, H. Y. Chang, W. J. Greenleaf, ATAC-seq: A method for assaying chromatin accessibility genome-wide. *Curr. Protoc. Mol. Biol.* **109**, 21.29.1–21.29.9 (2015). [doi:10.1002/0471142727.mb2129s109](https://doi.org/10.1002/0471142727.mb2129s109)
 121. P. Milani, R. Escalante-Chong, B. C. Shelley, N. L. Patel-Murray, X. Xin, M. Adam, B. Mandefro, D. Sareen, C. N. Svendsen, E. Fraenkel, Cell freezing protocol suitable for ATAC-Seq on motor neurons derived from human induced pluripotent stem cells. *Sci. Rep.* **6**, 25474 (2016). [doi:10.1038/srep25474](https://doi.org/10.1038/srep25474) [Medline](#)
 122. D. Jebb, Z. Huang, M. Pippel, G. M. Hughes, K. Lavrichenko, P. Devanna, S.

- Winkler, L. S. Jermini, E. C. Skirmuntt, A. Katzourakis, L. Burkitt-Gray, D. A. Ray, K. A. M. Sullivan, J. G. Roscito, B. M. Kirilenko, L. M. Dávalos, A. P. Corthals, M. L. Power, G. Jones, R. D. Ransome, D. K. N. Dechmann, A. G. Locatelli, S. J. Puechmaile, O. Fedrigo, E. D. Jarvis, M. Hiller, S. C. Vernes, E. W. Myers, E. C. Teeling, Six reference-quality genomes reveal evolution of bat adaptations. *Nature* **583**, 578–584 (2020). [doi:10.1038/s41586-020-2486-3](https://doi.org/10.1038/s41586-020-2486-3) [Medline](#)
123. Q. Li, J. B. Brown, H. Huang, P. J. Bickel, Measuring reproducibility of high-throughput experiments. *Ann. Appl. Stat.* **5**, 1752–1779 (2011). [doi:10.1214/11-AOAS466](https://doi.org/10.1214/11-AOAS466)
124. A. R. Quinlan, I. M. Hall, BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010). [doi:10.1093/bioinformatics/btq033](https://doi.org/10.1093/bioinformatics/btq033) [Medline](#)
125. Y. Liao, G. K. Smyth, W. Shi, featureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014). [doi:10.1093/bioinformatics/btt656](https://doi.org/10.1093/bioinformatics/btt656) [Medline](#)
126. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014). [doi:10.1186/s13059-014-0550-8](https://doi.org/10.1186/s13059-014-0550-8) [Medline](#)
127. H. Zhang, L. Song, X. Wang, H. Cheng, C. Wang, C. A. Meyer, T. Liu, M. Tang, S. Aluru, F. Yue, X. S. Liu, H. Li, Fast alignment and preprocessing of chromatin profiles with Chromap. *Nat. Commun.* **12**, 6566 (2021). [doi:10.1038/s41467-021-26865-w](https://doi.org/10.1038/s41467-021-26865-w) [Medline](#)
128. J. M. Granja, M. R. Corces, S. E. Pierce, S. T. Bagdatli, H. Choudhry, H. Y. Chang, W. J. Greenleaf, ArchR is a scalable software package for integrative single-cell chromatin accessibility analysis. *Nat. Genet.* **53**, 403–411 (2021). [doi:10.1038/s41588-021-00790-6](https://doi.org/10.1038/s41588-021-00790-6) [Medline](#)
129. Z. Yao, C. T. J. van Velthoven, T. N. Nguyen, J. Goldy, A. E. Sedeno-Cortes, F. Baftizadeh, D. Bertagnolli, T. Casper, M. Chiang, K. Crichton, S.-L. Ding, O. Fong, E. Garren, A. Glandon, N. W. Gouwens, J. Gray, L. T. Graybuck, M. J. Hawrylycz, D. Hirschstein, M. Kroll, K. Lathia, C. Lee, B. Levi, D. McMillen, S. Mok, T. Pham, Q. Ren, C. Rimorin, N. Shapovalova, J. Sulc, S. M. Sunkin, M. Tieu, A. Torkelson, H. Tung, K. Ward, N. Dee, K. A. Smith, B. Tasic, H. Zeng, A taxonomy of transcriptomic cell types across the isocortex and hippocampal formation. *Cell* **184**, 3222–3241.e26 (2021). [doi:10.1016/j.cell.2021.04.021](https://doi.org/10.1016/j.cell.2021.04.021) [Medline](#)
130. A. Frankish, M. Diekhans, A. M. Ferreira, R. Johnson, I. Jungreis, J. Loveland, J. M. Mudge, C. Sisu, J. Wright, J. Armstrong, I. Barnes, A. Berry, A. Bignell, S. Carbonell Sala, J. Chrast, F. Cunningham, T. Di Domenico, S. Donaldson, I. T. Fiddes, C. García Girón, J. M. Gonzalez, T. Grego, M. Hardy, T. Hourlier, T. Hunt, O. G. Izuogu, J. Lagarde, F. J. Martin, L. Martínez, S. Mohanan, P. Muir, F. C. P. Navarro, A. Parker, B. Pei, F. Pozo, M. Ruffier, B. M. Schmitt, E. Stapleton, M. M. Suner, I. Sycheva, B. Uszczyńska-Ratajczak, J. Xu, A. Yates, D. Zerbino, Y. Zhang, B. Aken, J. S. Choudhary, M. Gerstein, R. Guigó, T. J. P. Hubbard, M. Kellis, B. Paten, A. Raymond, M. L. Tress, P. Flicek, GENCODE reference annotation for the human and mouse genomes. *Nucleic Acids Res.* **47** (D1), D766–D773 (2019). [doi:10.1093/nar/gky955](https://doi.org/10.1093/nar/gky955) [Medline](#)
131. C. Y. McLean, D. Bristor, M. Hiller, S. L. Clarke, B. T. Schaar, C. B. Lowe, A. M. Wenger, G. Bejerano, GREAT improves functional interpretation of cis-regulatory regions. *Nat. Biotechnol.* **28**, 495–501 (2010). [doi:10.1038/nbt.1630](https://doi.org/10.1038/nbt.1630) [Medline](#)
132. U. Raudvere, L. Kolberg, I. Kuzmin, T. Arak, P. Adler, H. Peterson, J. Vilo, g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Res.* **47** (W1), W191–W198 (2019). [doi:10.1093/nar/gkz369](https://doi.org/10.1093/nar/gkz369) [Medline](#)
133. E. Kulinskaya, On two-sided p-values for non-symmetric distributions. [arXiv:0810.2124](https://arxiv.org/abs/0810.2124) [math.ST] (2008).
134. Y. Wang, X. Zhang, F. Chen, L. Chen, J. Wang, J. Xie, LRRK2-NFATc2 Pathway Associated with Neuroinflammation May Be a Potential Therapeutic Target for Parkinson’s Disease. *J. Inflamm. Res.* **14**, 2583–2586 (2021). [doi:10.2147/JIR.S301531](https://doi.org/10.2147/JIR.S301531) [Medline](#)
135. P. V. Lovell, N. A. Huizinga, S. R. Friedrich, M. Wirthlin, C. V. Mello, The constitutive differential transcriptome of a brain circuit for vocal learning. *BMC Genomics* **19**, 231 (2018). [doi:10.1186/s12864-018-4578-0](https://doi.org/10.1186/s12864-018-4578-0) [Medline](#)
136. B. St Pourcain, K. Wang, J. T. Glessner, J. Golding, C. Steer, S. M. Ring, D. H. Skuse, S. F. A. Grant, H. Hakonarson, G. Davey Smith, Association between a high-risk autism locus on 5p14 and social communication spectrum phenotypes in the general population. *Am. J. Psychiatry* **167**, 1364–1372 (2010). [doi:10.1176/appi.ajp.2010.09121789](https://doi.org/10.1176/appi.ajp.2010.09121789) [Medline](#)
137. D. A. Sweetser, I. Elsharkawi, L. Yonker, M. Steeves, K. Parkin, R. Thibert, “Pitt-Hopkins Syndrome” in *GeneReviews*, M. P. Adam, H. H. Ardinger, R. A. Pagon, S. E. Wallace, L. J. H. Bean, G. Mirzaa, A. Amemiya, Eds. (University of Washington, 2012).
138. G. Gat-Yablonski, R. Frumkin-Ben David, M. Bar, O. Potievsky, M. Phillip, L. Lazar, Homozygous microdeletion of the POU1F1, CHMP2B, and VGLL3 genes in chromosome 3—A novel syndrome. *Am. J. Med. Genet. A.* **155**, 2242–2246 (2011). [doi:10.1002/ajmg.a.34136](https://doi.org/10.1002/ajmg.a.34136) [Medline](#)
139. O. Bartsch, D. Schindler, V. Beyer, S. Gesk, R. van’t Slot, I. Feddersen, A. Buijs, N. G. J. Jaspers, R. Siebert, T. Haaf, M. Poot, A girl with an atypical form of ataxia telangiectasia and an additional de novo 3.14 Mb microduplication in region 19q12. *Eur. J. Med. Genet.* **55**, 49–55 (2012). [doi:10.1016/j.ejmg.2011.08.001](https://doi.org/10.1016/j.ejmg.2011.08.001) [Medline](#)
140. S. Chowdhury, A. M. Bandholz, S. Parkash, S. Dyack, A. L. Rideout, K. A. Leppig, H. Thiese, P. G. Wheeler, M. Tsang, B. C. Ballif, L. G. Shaffer, B. S. Torchia, J. W. Ellison, J. A. Rosenfeld, Phenotypic and molecular characterization of 19q12q13.1 deletions: A report of five patients. *Am. J. Med. Genet. A.* **164**, 62–69 (2014). [doi:10.1002/ajmg.a.36201](https://doi.org/10.1002/ajmg.a.36201) [Medline](#)
141. R. Nacinovich, N. Villa, F. Broggi, C. Tavaniello, M. Bomba, D. Conconi, S. Redaelli, E. Sala, M. Lavitrano, F. Neri, 19q12q13.2 duplication syndrome: Neuropsychiatric long-term follow-up of a new case and literature update. *Neuropsychiatr. Dis. Treat.* **13**, 2545–2550 (2017). [doi:10.2147/NDT.S142356](https://doi.org/10.2147/NDT.S142356) [Medline](#)
142. V. Murcia Pienkowski, M. Kucharczyk, M. Młynek, K. Szczałuba, M. Rydzanicz, B. Poszewiecka, A. Skórka, M. Sykulski, A. Biernacka, A. A. Koppolu, R. Posmyk, A. Walczak, J. Kosińska, P. Krajewski, J. Castaneda, E. Obersztyn, E. Jurkiewicz, R. Śmigiel, A. Gambin, K. Chrzanowska, M. Krajewska-Walasek, R. Płoski, Mapping of breakpoints in balanced chromosomal translocations by shallow whole-genome sequencing points to *EFNA5*, *BAHD1* and *PPP2R5E* as novel candidates for genes causing human Mendelian disorders. *J. Med. Genet.* **56**, 104–112 (2019). [doi:10.1136/jmedgenet-2018-105527](https://doi.org/10.1136/jmedgenet-2018-105527) [Medline](#)
143. S. J. Kraft, thesis, University of Illinois at Urbana-Champaign (2010).
144. A. T. Pagnamenta, H. Khan, S. Walker, D. Gerrelli, K. Wing, M. C. Bonaglia, R. Giorda, T. Berney, E. Mani, M. Molteni, D. Pinto, A. Le Couteur, J. Hallmayer, J. S. Sutcliffe, P. Szatmari, A. D. Paterson, S. W. Scherer, V. J. Veland, A. P. Monaco, Rare familial 16q21 microdeletions under a linkage peak implicate cadherin 8 (*CDH8*) in susceptibility to autism and learning disability. *J. Med. Genet.* **48**, 48–54 (2011). [doi:10.1136/jmg.2010.079426](https://doi.org/10.1136/jmg.2010.079426) [Medline](#)
145. A. Karoglan, D. Schanze, C. Bär, P. Muschke, M. Zenker, I. Schanze, A 2q24.2 microdeletion containing *TANK* as novel candidate gene for intellectual disability. *Am. J. Med. Genet. A.* **179**, 832–836 (2019). [doi:10.1002/ajmg.a.61093](https://doi.org/10.1002/ajmg.a.61093) [Medline](#)
146. P. Deriziotis, B. J. O’Roak, S. A. Graham, S. B. Estruch, D. Dimitropoulou, R. A. Bernier, J. Gerdtts, J. Shendure, E. E. Eichler, S. E. Fisher, De novo TBR1 mutations in sporadic autism disrupt protein functions. *Nat. Commun.* **5**, 4954 (2014). [doi:10.1038/ncomms5954](https://doi.org/10.1038/ncomms5954) [Medline](#)
147. O. Palumbo, M. Fichera, P. Palumbo, R. Rizzo, E. Mazzolla, D. M. Cocuzza, M. Carella, T. Mattina, TBR1 is the candidate gene for intellectual disability in patients with a 2q24.2 interstitial deletion. *Am. J. Med. Genet. A.* **164**, 828–833 (2014). [doi:10.1002/ajmg.a.36363](https://doi.org/10.1002/ajmg.a.36363) [Medline](#)
148. J. J. Orsini, M. L. Escobar, M. P. Wasserstein, M. Caggana, “Krabbe Disease” in *GeneReviews*, M. P. Adam, H. H. Ardinger, R. A. Pagon, S. E. Wallace, L. J. H. Bean, G. Mirzaa, A. Amemiya, Eds. (Univ. of Washington, 2000).
149. S. Schlaubitz, S. A. Yatsenko, L. D. Smith, K. L. Keller, L. E. Vissers, D. A. Scott, W. W. Cai, W. Reardon, O. A. Abdul-Rahman, E. J. Lammer, C. A. Lifchez, E. Magenis, J. A. Veltman, P. Stankiewicz, B. U. Zabel, B. Lee, Ovotestes and XY sex reversal in a female with an interstitial 9q33.3-q34.1 deletion encompassing *NR5A1* and *LMX1B* causing features of Genitopatellar syndrome. *Am. J. Med. Genet. A.* **143A**, 1071–1081 (2007). [doi:10.1002/ajmg.a.31685](https://doi.org/10.1002/ajmg.a.31685) [Medline](#)
150. V. W. Hu, A. Addington, A. Hyman, Novel autism subtype-dependent genetic

- variants are revealed by quantitative trait and subphenotype association analyses of published GWAS data. *PLoS ONE* **6**, e19067 (2011). [doi:10.1371/journal.pone.0019067](https://doi.org/10.1371/journal.pone.0019067) [Medline](#)
151. S. Dimassi, A. Labalme, G. Lesca, G. Rudolf, N. Bruneau, E. Hirsch, A. Arzimanoglou, J. Motte, A. de Saint Martin, N. Boutry-Kryza, R. Cloarec, A. Benitto, A. Ameil, P. Edery, P. Rylvlin, J. De Bellescize, P. Szepietowski, D. Sanlaville, A subset of genomic alterations detected in rolandic epilepsies contains candidate or known epilepsy genes including GRIN2A and PRRT2. *Epilepsia* **55**, 370–378 (2014). [doi:10.1111/epi.12502](https://doi.org/10.1111/epi.12502) [Medline](#)
152. A. Lehman, S. Thouta, G. M. S. Mancini, S. Naidu, M. van Slegtenhorst, K. McWalter, R. Person, J. Mwenifumbo, R. Salvarinova, I. Guella, M. B. McKenzie, A. Datta, M. B. Connolly, S. M. Kalkhoran, D. Poburko, J. M. Friedman, M. J. Farrer, M. Demos, S. Desai, T. Claydon, S. Adam, C. du Souich, A. M. Elliott, A. Lehman, J. Mwenifumbo, T. N. Nelson, C. van Karnebeek, J. M. Friedman, S. Adam, C. Boelman, C. Bolbocean, S. E. Buerki, T. Candido, P. Eydoux, D. M. Evans, W. Gibson, G. Horvath, L. Huh, T. N. Nelson, G. Sinclair, T. Tarling, E. B. Toyota, K. N. Townsend, M. I. Van Allen, C. van Karnebeek, S. Vercauteren; CAUSES Study; EPGEN Study, Loss-of-Function and Gain-of-Function Mutations in KCNQ5 Cause Intellectual Disability or Epileptic Encephalopathy. *Am. J. Hum. Genet.* **101**, 65–74 (2017). [doi:10.1016/j.ajhg.2017.05.016](https://doi.org/10.1016/j.ajhg.2017.05.016) [Medline](#)
153. K. J. Franek, J. Butler, J. Johnson, R. Simensen, M. J. Friez, F. Bartel, T. Moss, B. DuPont, K. Berry, M. Bauman, C. Skinner, R. E. Stevenson, C. E. Schwartz, Deletion of the immunoglobulin domain of IL1RAPL1 results in nonsyndromic X-linked intellectual disability associated with behavioral problems and mild dysmorphism. *Am. J. Med. Genet. A.* **155**, 1109–1114 (2011). [doi:10.1002/ajmg.a.33833](https://doi.org/10.1002/ajmg.a.33833) [Medline](#)
154. D. Lehalle, D. Sanlaville, A. Guimier, E. Plouvier, T. Leblanc, L. Galmiche, I. Radford, S. Romana, L. Colleaux, L. de Pontual, S. Lyonnet, J. Amiel, Multiple congenital anomalies-intellectual disability (MCA-ID) and neuroblastoma in a patient harboring a de novo 14q23.1q23.3 deletion. *Am. J. Med. Genet. A.* **164**, 1310–1317 (2014). [doi:10.1002/ajmg.a.36452](https://doi.org/10.1002/ajmg.a.36452) [Medline](#)
155. Y. Le Guen, F. Leroy, C. Philippe, J.-F. Mangin, G. Dehaene-Lambertz, V. Frouin; IMAGEN Consortium, Enhancer Locus in ch14q23.1 Modulates Brain Asymmetric Temporal Regions Involved in Language Processing. *Cereb. Cortex* **30**, 5322–5332 (2020). [doi:10.1093/cercor/bhaa112](https://doi.org/10.1093/cercor/bhaa112) [Medline](#)
156. G. Lesca, G. Rudolf, A. Labalme, E. Hirsch, A. Arzimanoglou, P. Genton, J. Motte, A. de Saint Martin, M.-P. Valenti, C. Boulay, J. De Bellescize, P. Kéo-Kosal, N. Boutry-Kryza, P. Edery, D. Sanlaville, P. Szepietowski, Epileptic encephalopathies of the Landau-Kleffner and continuous spike and waves during slow-wave sleep types: Genomic dissection makes the link with autism. *Epilepsia* **53**, 1526–1538 (2012). [doi:10.1111/j.1528-1167.2012.03559.x](https://doi.org/10.1111/j.1528-1167.2012.03559.x) [Medline](#)
157. H. Y. Zoghbi, H. T. Orr, Pathogenic mechanisms of a polyglutamine-mediated neurodegenerative disease, spinocerebellar ataxia type 1. *J. Biol. Chem.* **284**, 7425–7429 (2009). [doi:10.1074/jbc.R800041200](https://doi.org/10.1074/jbc.R800041200) [Medline](#)
158. P. B. Celestino-Soper, C. Skinner, R. Schroer, P. Eng, J. Shenai, M. M. Nowaczyk, D. Terespolsky, D. Cushing, G. S. Patel, L. Immken, A. Willis, J. Wiszniewska, R. Matalon, J. A. Rosenfeld, R. E. Stevenson, S.-H. L. Kang, S. W. Cheung, A. L. Beaudet, P. Stankiewicz, Deletions in chromosome 6p22.3-p24.3, including ATXN1, are associated with developmental delay and autism spectrum disorders. *Mol. Cytogenet.* **5**, 17 (2012). [doi:10.1186/1755-8166-5-17](https://doi.org/10.1186/1755-8166-5-17) [Medline](#)
159. P. Benckek, R. P. Igo, H. Voss-Hoynes, Y. Wren, G. Miller, B. Truitt, W. Zhang, M. Osterman, L. Freebairn, J. Tag, H. G. Taylor, E. R. Chan, P. Roussos, B. Lewis, C. M. Stein, S. K. Iyengar, Association between genes regulating neural pathways for quantitative traits of speech and language disorders. *npj. Genomic Med.* **6**, 1–11 (2021).
160. D. Milani, C. Sabatini, F. M. P. Manzoni, P. F. Ajmone, C. Rigamonti, M. Malacarne, M. Pierluigi, S. Cavani, M. A. Costantino, Microdeletion 2q23.3q24.1: Exploring genotype-phenotype correlations. *Congenit. Anom. (Kyoto)* **55**, 107–111 (2015). [doi:10.1111/cga.12080](https://doi.org/10.1111/cga.12080) [Medline](#)
161. C. Benito-Aragón, R. Gonzalez-Sarmiento, T. Liddell, I. Diez, F. d’Oleire Uquillas, L. Ortiz-Terán, E. Bueichekú, H. M. Chow, S.-E. Chang, J. Sepulcre, Neurofilament-lysosomal genetic intersections in the cortical network of stuttering. *Prog. Neurobiol.* **184**, 101718 (2020). [doi:10.1016/j.pneurobio.2019.101718](https://doi.org/10.1016/j.pneurobio.2019.101718) [Medline](#)
162. S. Kielb, Y. Y. Kisanuki, E. Dawson, Neuropsychological profile associated with an alpha-synuclein gene (SNCA) duplication. *Clin. Neuropsychol.* **36**, 1–12 (2021). [Medline](#)

ACKNOWLEDGMENTS

We would like to thank Andrew C. Halley and Leah Krubitzer for helpful discussions and critical assistance in identifying orofacial and wing subregions of the *Rousettus aegyptiacus* motor cortex. We would like to thank Frederic E. Theunissen for helpful discussions around the analysis of coherence of neural activity and for sharing his lab’s code. We would like to thank the other members of the Zoonomia Consortium, the Vertebrate Genomes Project Vocal Learning working group, and the Pfenning, Yartsev, and Jarvis labs for useful discussions and suggestions. **Funding:** Alfred P. Sloan Foundation Research Fellowship (ARP); National Science Foundation grant NSF 20-525 (ARP); National Institutes of Health grant NIDA DP1DA046585 (ARP); National Institutes of Health grant NIDA F30DA053020 (BNP); National Institutes of Health grant R01NS11479 (EA); National Institutes of Health grant R01 NS121231 (BL); National Science Foundation grant NSF IOS-2022241 (BL); National Institutes of Health grant DP2 DP2-DC016164 (MMY); The New York Stem Cell Foundation NYSCF-R-NI40 (MMY); Alfred Sloan Foundation FG-2017-9646 (MMY); Brain Research Foundation BRFSG-2017-09 (MMY); Packard Foundation Fellowship 2017-66825 (MMY); Klingenstein-Simons Fellowship (MMY); Human Frontiers Research Program (MMY); Pew Charitable trust 00029645 (MMY); McKnight Foundation 042823 (MMY); Carnegie Mellon University Computational Biology Department Lane Fellowship (IMK); Dana Foundation (MMY). This work used the Extreme Science and Engineering Discovery Environment (XSEDE), through the Pittsburgh Supercomputing Center Bridges and Bridges-2 Compute Clusters, which was supported by the National Science Foundation (grant TG-BIO200055). Portions of this research were conducted on Lehigh University’s Research Computing infrastructure partially supported by NSF Award 2019035 (WKM). National Institutes of Health grant 2R01 HG009299-06A1 (MC, NLC, SLKP); National Science Foundation grant 2238125 (MC). **Author Contributions:** Conceptualization: ARP, MEW, TAS, JEE, MMY; Methodology: ARP, MEW, TAS, JEE, MMY, IMK; Resources: TAS, MMY, ARB, WM, EA, BKL; Data Curation: TAS, JEE, XZ, BNP, AZW; Investigation: MEW, TAS, JEE, VAS, AR, MBJ, NSB, TL, MMY, ARP; Software: MEW, XZ, IMK, DES, SA, JEE, APR; Formal analysis: MEW, XZ, IMK, DES, AJL, ARP, MMY, TAS, JEE, VAS, AR, MBJ, NSB; Funding acquisition: ARP, MMY; Supervision: ARP, MEW, MMY, TAS, IMK; Writing – original draft: MEW, JEE, TAS, ARP; Writing – review and editing: MEW, JEE, TAS, IMK, DES, AJL, MMY, ARP. **Competing interests:** ARP is founder and CEO of Snail Biosciences. **Data and materials availability:** Egyptian fruit bat ATAC-seq data—including raw .fasta files, .bigWig files of genome-aligned reads, processed open chromatin .narrowpeak files, and extensive sample metadata—have been made available at GEO under accession ID GSE187366. Publicly available ATAC-Seq data was obtained from GEO GSE159815 and GSE161374. Single-nucleus open chromatin was downloaded from the NEMO archive: https://data.nemoarchive.org/biccn/grant/u19_cemba/cemba/epigenome/scell/ATACseq/mouse/. The code for the electrophysiology and functional components of the project are available here: <https://github.com/NeuroBatLab/LoggerDataProcessing/>; <https://github.com/NeuroBatLab/SoundAnalysisBats/>; <https://github.com/NeuroBatLab/LoggerDataProcessing/>; https://github.com/julieelie/Kilosort2_Tetrode/configFiles/configFile16.m. The code for the convergent evolution analysis can be found at <https://github.com/pfenninglab/TACIT>, <https://github.com/veg/hyphy> and <https://github.com/veg/hyphy-analyses/tree/master/BUSTED-PH>. The histological and electrophysiology source data from this study are available on FigShare under the DOI 10.6084/m9.figshare.25180430. The code for conducting and visualizing the RERconverge analysis can be found here:

<https://zenodo.org/doi/10.5281/zenodo.10641176> (95). The code for conducting and visualizing the TACIT analysis can be found here: <https://zenodo.org/doi/10.5281/zenodo.5952292> (96). **License information:** Copyright © 2024 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. <https://www.science.org/about/science-licenses-journal-article-reuse>

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abn3263

Materials and Methods

Supplementary Text

Figures S1 to S7

Tables S1 and S2

References (98–162)

MDAR Reproducibility Checklist

Movie S1

Data S1 to S10

Accepted 20 February 2024

Published online 29 February 2024

10.1126/science.abn3263

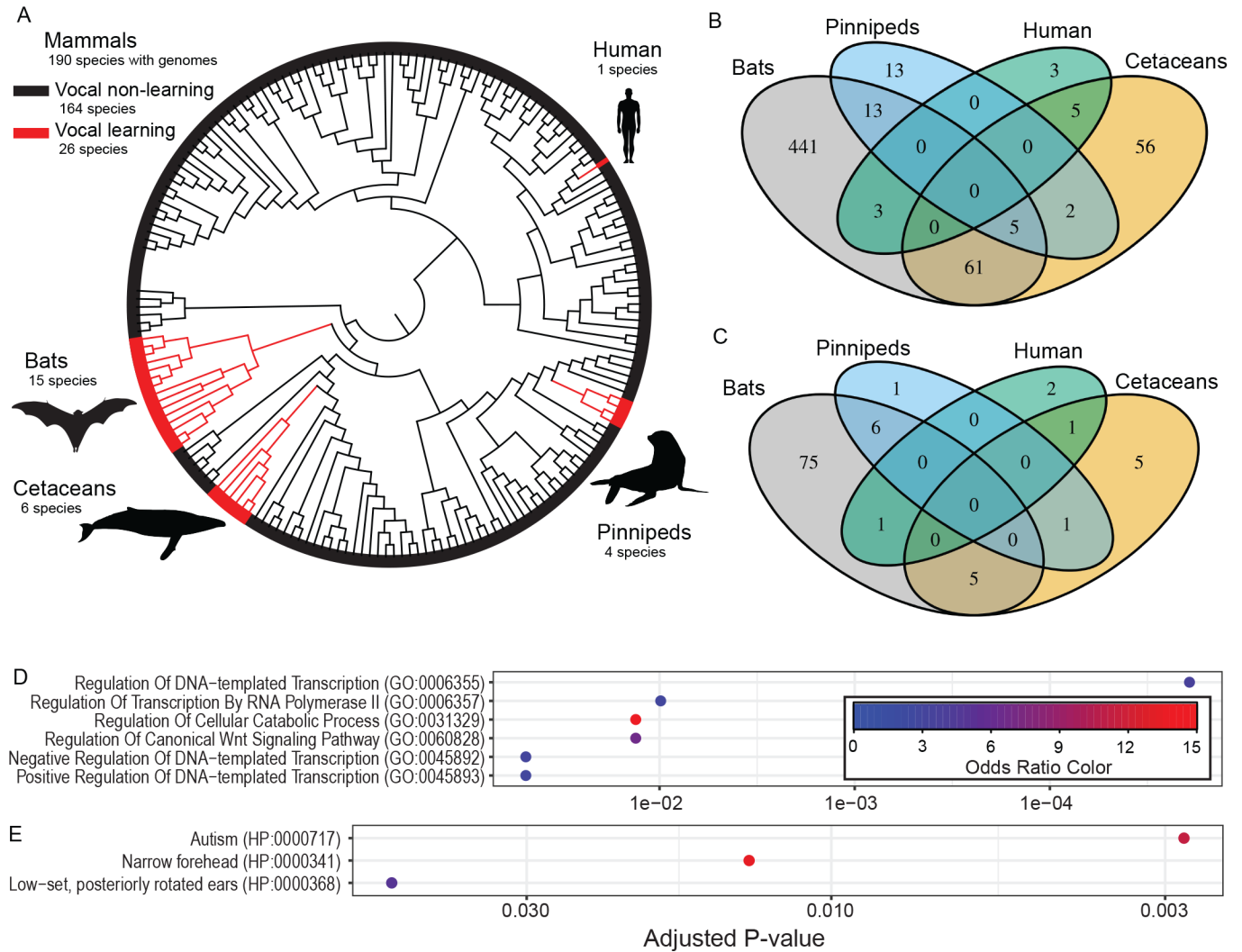


Fig. 1. Convergent changes in protein sequence associated with vocal learning evolution across 215 mammalian species. (A) A cladogram of mammalian species whose genomes were analyzed in this study highlights the convergent evolution of vocal learning species (in red) relative to non-learners (in black). The phylogenetic tree used in our analyses was derived from (97). Each of the genes implicated by RERconverge with lower (B) or higher (C) evolutionary rates in vocal learners are annotated based on whether or not they show a significant signature within the four vocal learning clades based on a Bayes factor ≥ 5 (18). All significant gene ontology categories (adjusted $p < 0.10$, EnrichR) are plotted for the 200 genes with conserved (D) and accelerated (E) selection in vocal learning clades, based on the combination of RERconverge and HyPhy RELAX. The points are colored by the odds ratio within the set of implicated genes relative to the genes outside of the set, which corresponds to the degree of enrichment within that set.

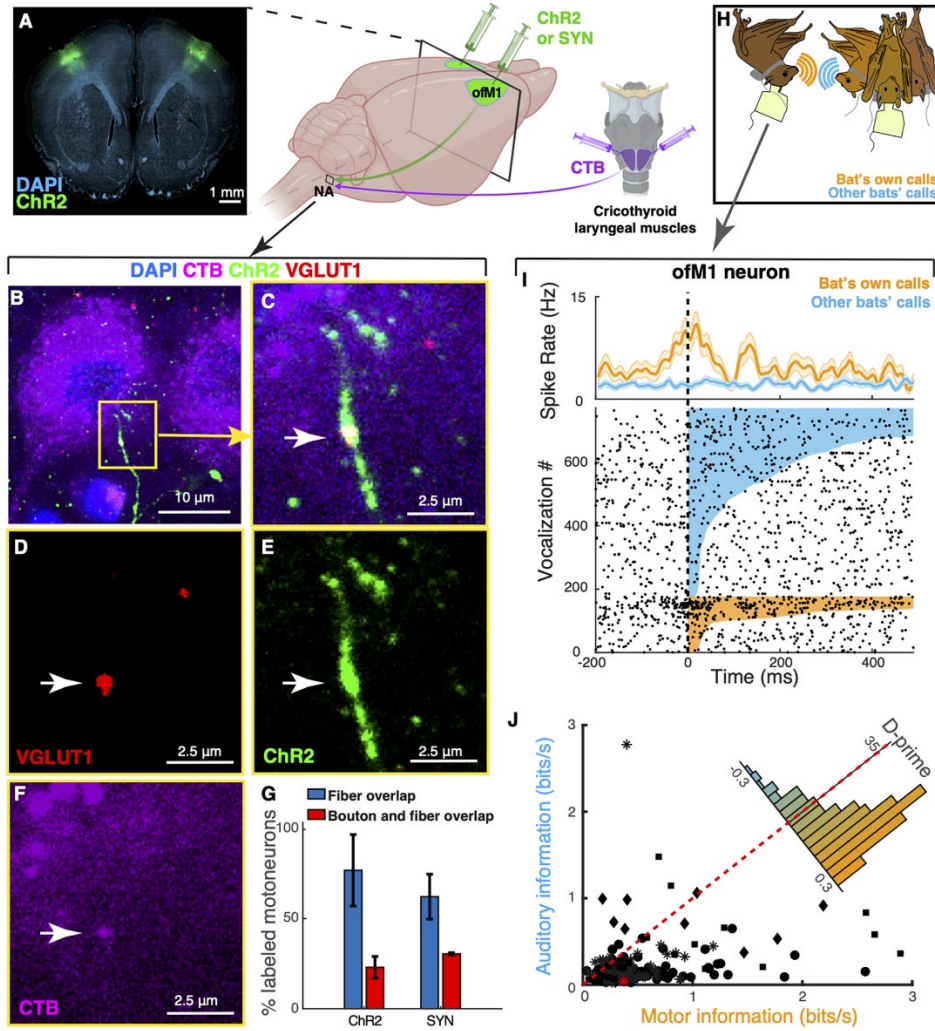


Fig. 2. Identification of an anatomically specialized motor cortical region targeting laryngeal motoneurons in the Egyptian fruit bat. (A) Right: schematic of anatomical tracing approaches. Retrograde tracer cholera toxin B (CTB, purple) was injected bilaterally into the cricothyroid muscles to label brainstem motoneurons in nucleus ambiguus (NA). Simultaneously, an anterograde viral tracer (channelrhodopsin-2, ChR2, or Synapsin/synaptophysin dual-label, SYN; green) was injected bilaterally into the orofacial motor cortex (ofM1) to label corticobulbar projections into NA. Left: example coronal section showing cortical injection sites with anterograde tracer (ChR2, green) and DAPI labeling (cyan). (B to F) Laryngeal motoneurons in the NA identified using a retrograde tracer (CTB, purple), cortical fibers labeled with ChR2 (green), corticobulbar synapses labeled with VGLUT1 (red), and DAPI (blue). (B) and (C) are overlaid images showing colocalization of fibers with a synaptic bouton on the retrograde labeled cell (white arrow). (G) Percentage of laryngeal motoneurons labeled with CTB that are colocalized with cortical fibers (blue) or with both cortical fibers and synaptic boutons (red). Note that both tracing techniques qualitatively yielded similar results: ChR2, $n = 51$ cells from 3 bats; Synapsin/synaptophysin dual-label virus (SYN), $n = 26$ cells from 2 bats. (H) Illustration of the experimental setup during which wireless electrophysiological recordings were conducted from the identified cortical region in freely behaving and vocalizing bats. (I) Spiking activity of an example ofM1 neuron aligned to the onset of vocalizations produced (bat's own calls, orange) or heard (other bats' calls, blue) by the bat subject. *Top row*, time varying mean firing rate and corresponding raster plot below. Colored lines in the raster plot show the duration of each vocalization. Note the increased firing rate during vocal production as compared to hearing. (J) Information (see Methods) between the time varying firing rate and the amplitude of produced (x-axis) vs. heard (y-axis) vocalizations for 219 single units (marker shapes indicate bat ID, $n=4$ bats). The cell shown in (I) is highlighted in red. Inset shows the distribution of D-prime between motor and auditory information for the same cells. Note that the distribution is heavily skewed toward higher motor information rather than auditory information coded in the activity of the recorded neurons. Error bars are mean \pm SEM throughout the figure.

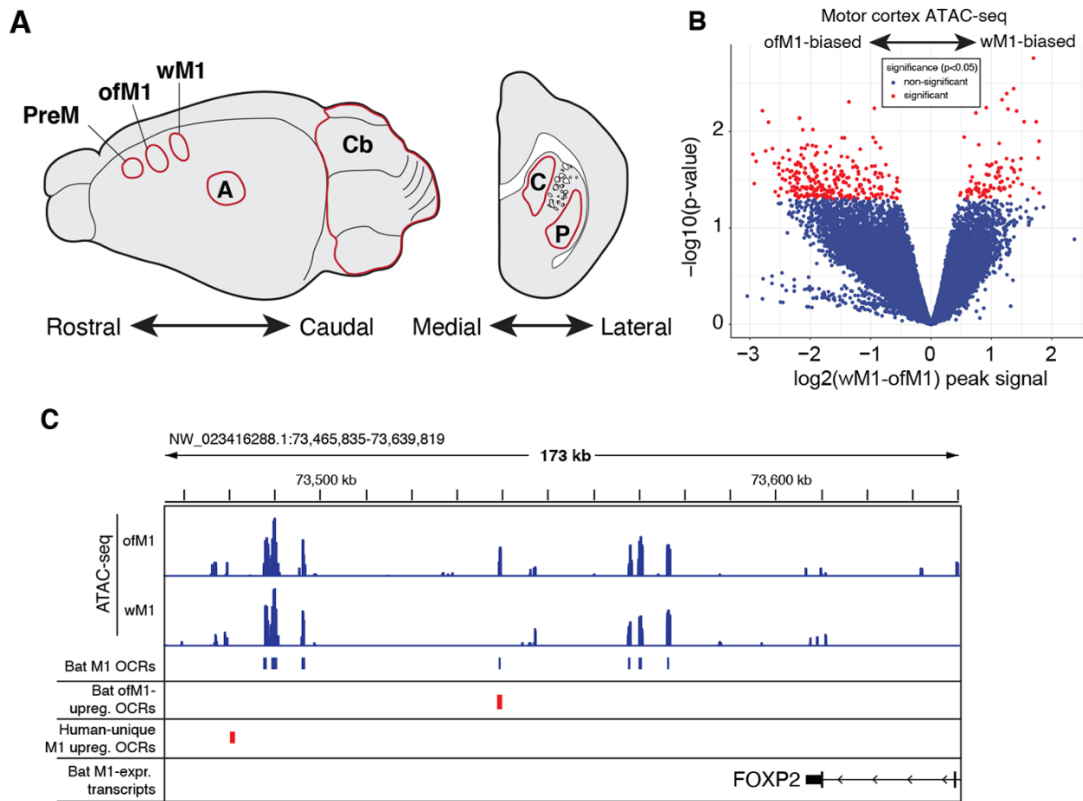


Fig. 3. Differential Open Chromatin in Bat Orofacial M1 relative to Wing M1. (A) Open chromatin was profiled from 7 dissected brain regions of Egyptian Fruit bats. (B) Volcano plot of ATAC-seq OCRs with differential activity between the orofacial and wing subregions of primary motor cortex (ofM1 and wM1, respectively) of Egyptian fruit bat. (C) Genome browser showing ofM1 and wM1 ATAC-seq traces at the 3' end of the *FOXP2* locus. Reproducible M1 open chromatin regions (OCRs) are indicated in blue, with a differentially active OCR in ofM1 relative to wM1 highlighted in red.

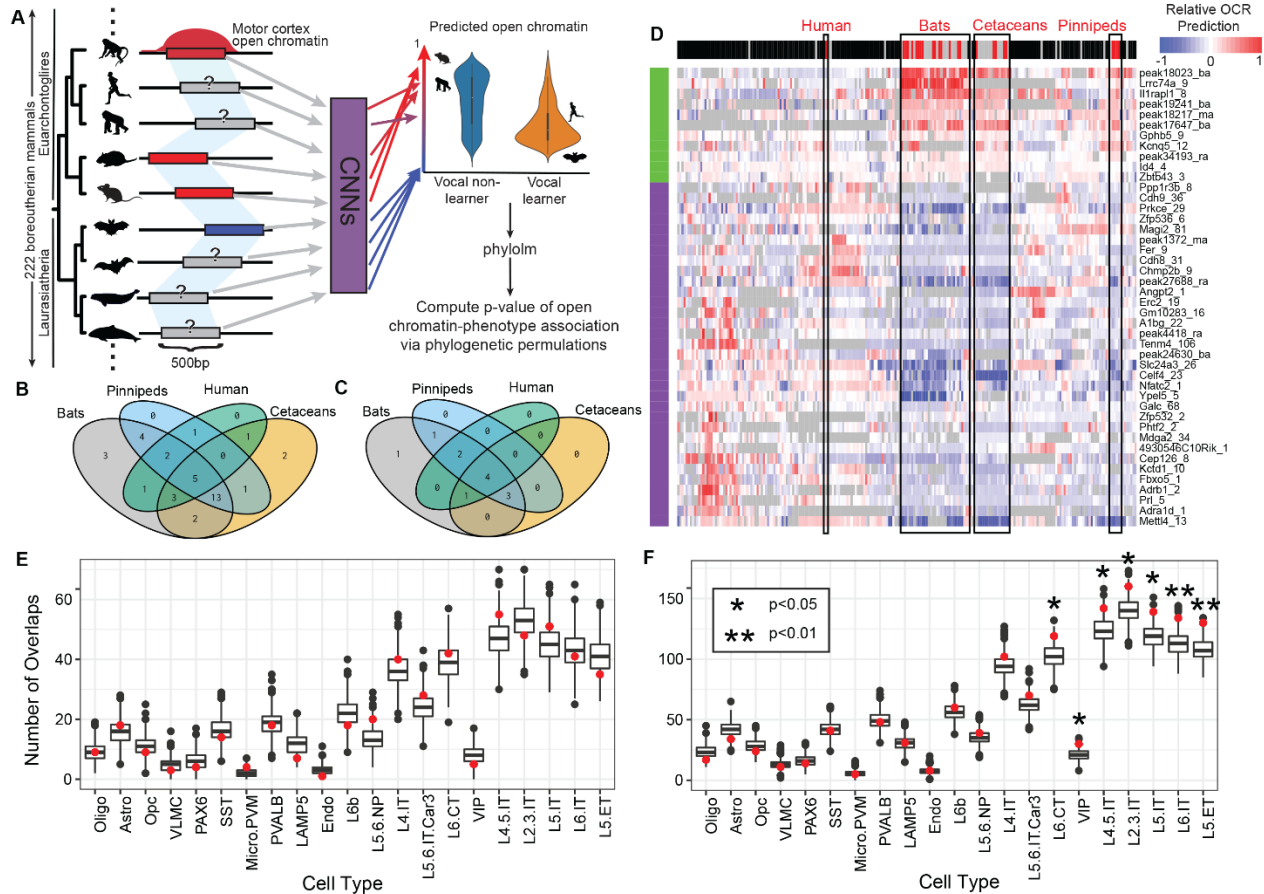


Fig. 4. Vocal learning-associated convergent evolution in motor cortex open chromatin regions implicates specific neuron subtypes. (A) Overview of applying the Tissue-Aware Conservation Inference Toolkit [TACIT (23)] approach to vocal learning. OCRs (left) identified in motor cortex (M1). Measured open chromatin from M1 (4 species) were used to train convolutional neural networks (CNNs) to predict M1 open chromatin from sequence alone. Red bars and corresponding arrows indicate the presence of a peak while the blue bars represent the absence. The same OCRs were then mapped across 222 mammalian genomes (left) and the identified sequences were used as input to the CNNs to predict open chromatin activity. TACIT identified OCRs whose predicted open chromatin across species was significantly associated with those species' vocal learning status. (B and C) The 4-way Venn diagrams represent the number of OCRs implicated by TACIT (both M1 and PV+) as displaying low (B) or high (C) activity in each of the vocal learning clades based on a *t* test. (D) The heatmap visualizes specific open chromatin regions along the rows (predicted higher in vocal learners in green; predicted lower in vocal learners in purple) across 222 mammals in the columns (vocal learner in red, vocal non-learner in black, insufficient or conflicting evidence in gray). The color in each cell corresponds to the z-scored predicted open chromatin, with low open chromatin in blue, mean open chromatin in white, and high open chromatin in red. For open chromatin regions predicted to be significantly less (E) or more (F) open in vocal learning species ($p < 0.05$), the red point shows the number of overlapping regions (y-axis) across mouse cortical cell types (x-axis). The bar-plot shows the distribution across 1,000 permutations of the peaks implicated by TACIT. The notches extend $1.58 * IQR / \sqrt{n}$, which gives a roughly 95% confidence).