

RESEARCH ARTICLE SUMMARY

BRAIN EVOLUTION

Enhancer-driven cell type comparison reveals similarities between the mammalian and bird pallium

Nikolai Hecker[†], Niklas Kempynck[†], David Mauduit, Darina Abaffyová, Roel Vandepoel, Sam Dieltiens, Lars Borm, Ioannis Sarropoulos, Carmen Bravo González-Blas, Julie De Man, Kristofer Davie, Elke Leysen, Jeroen Vandensteen, Rani Moors, Gert Hulselmans, Lynette Lim, Joris De Wit, Valerie Christiaens, Suresh Poovathingal, Stein Aerts^{*}

INTRODUCTION: The identity of cell types is governed by gene regulatory networks, which comprise cell type-specific combinations of transcription factors (TFs) that bind to genomic enhancer regions. The arrangements of TF binding sites form cell type-specific enhancer codes. Deep learning models trained on single-cell data provide the means to model and characterize enhancer codes at nucleotide resolution. Enhancer codes at such resolution have not yet been characterized for the mammalian telencephalon, which constitutes a major part of the forebrain, including the pallium. The pallium displays notable neuroanatomical differences between mammals and nonmammalian vertebrates. Most noticeable, the mammalian pallium contains a six-layered neocortex that is absent in all nonmammalian

vertebrates, such as birds. Homologies between the mammalian and bird pallium are subject to a decades-long debate. It is currently unknown whether enhancer codes are conserved across vertebrate brains and whether they are informative to resolve homology relationships between species at the cell type level.

RATIONALE: To characterize and compare the enhancer codes of brain cell types between the mammalian and bird pallium, we generated single-cell multiome (scMultiome) and spatially resolved transcriptomes data of the chicken telencephalon. As a baseline to map cell type similarities between species, we compared the transcriptomes of telencephalon cell types between human, mouse, and chicken. We then used variable chromatin accessibility as a proxy

to identify potential genomic enhancer regions and to assess their cell type specificity. Next, we trained sequence-based deep learning models on these regions to infer cell type-specific enhancer codes for the human, mouse, and chicken telencephalon. We implemented three metrics that exploit enhancer codes to compare cell types between species.

RESULTS: Excitatory neurons of the chicken telencephalon distinctly localize to pallial neuroanatomical regions, including the mesopallium, entopallium, hyperpallium, and nidopallium. Based on the transcriptomic and enhancer code comparisons, nonneuronal and γ -aminobutyric acid-mediated (GABAergic) cell types show a high degree of similarity across birds and mammals, which is reflected by conserved TF combinations for these cell types. On the other hand, the enhancer codes of excitatory neurons in the mammalian and avian pallium exhibit a higher degree of divergence. These matches only partially agree with existing evolutionary models for homologies between vertebrate pallial cell types based on developmental trajectories and brain circuitry. We found that the mammalian deep-layer excitatory neurons are most similar to mesopallial neurons, and mammalian neocortical upper-layer, piriform cortex, and amygdalar neurons are most similar to hyper- and nidopallial neurons. As a validation for the predicted correspondences between mammalian and bird cell types, we performed *in vivo* enhancer reporter assays. We show that chicken enhancer sequences exhibit activity in the corresponding mammalian telencephalic cell types when assayed in mouse brains.

CONCLUSION: Our study shows that enhancer codes can be exploited to infer cell type correspondences between species that are in line with transcriptomic comparisons. Joint comparisons of transcriptomes and deep learning-based enhancer codes reveal both expected and unexpected correspondences between cell types in the mammalian and avian telencephalon, indicating conserved regulatory programs that likely originated in the common amniote ancestor and have been co-opted or diversified. The proposed enhancer code-based approaches are generally applicable and can be used to characterize and compare cell types across species using the genomic regulatory code. ■

The list of author affiliations is available in the full article online.

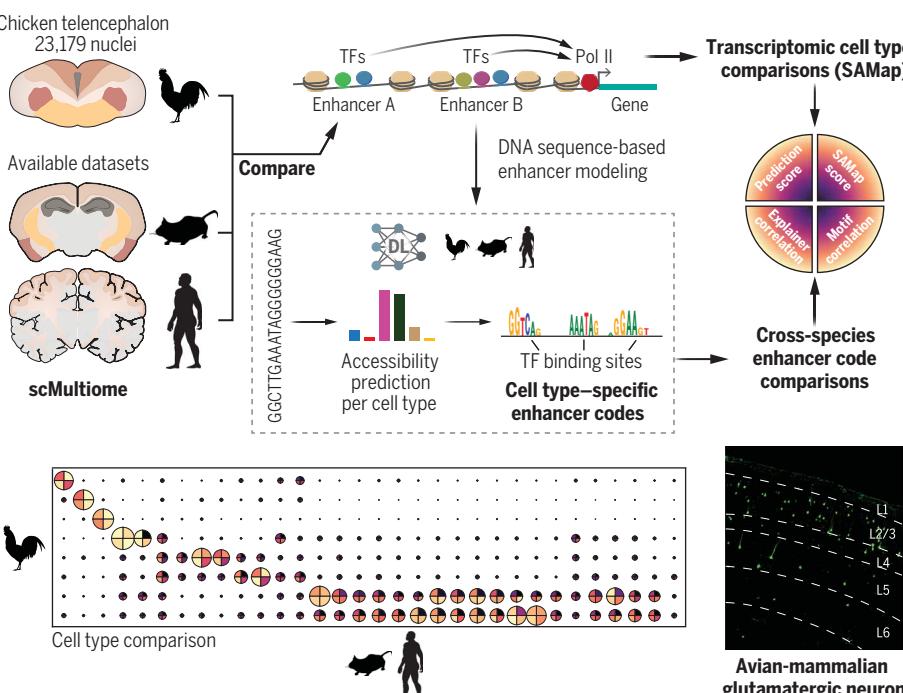
*Corresponding author. Email: stein.aerts@kuleuven.be

[†]These authors contributed equally to this work.

Cite this article as N. Hecker et al., *Science* 387, eadp3957 (2025). DOI: 10.1126/science.adp3957

S READ THE FULL ARTICLE AT

<https://doi.org/10.1126/science.adp3957>



Correspondences between avian and mammalian telencephalic regulatory codes hint at ancestral conservation. We generated a scMultiome and spatial chicken telencephalon dataset and compared chicken, mouse, and human telencephalic cell types by comparing their transcriptomic profiles and their inherent enhancer codes through sequence-based deep learning models. Pol II, polymerase II.

RESEARCH ARTICLE

BRAIN EVOLUTION

Enhancer-driven cell type comparison reveals similarities between the mammalian and bird pallium

Nikolai Hecker^{1,2,3†}, Niklas Kempynck^{1,2,3†}, David Mauduit^{1,2,3}, Darina Abaffyová^{1,2,3}, Roel Vandepoel^{1,2,3}, Sam Dieltiens^{1,2,3}, Lars Borm^{1,2,3}, Ioannis Sarropoulos⁴, Carmen Bravo González-Blas^{2,3}, Julie De Man^{1,2,3}, Kristofer Davie^{2,5}, Elke Leysen^{2,5}, Jeroen Vandensteen^{2,5}, Rani Moors^{2,5}, Gert Hulselmans^{1,2,3}, Lynette Lim^{2,5}, Joris De Wit^{2,5}, Valerie Christiaens^{1,2,3}, Suresh Poovathingal², Stein Aerts^{1,2,3*}

Combinations of transcription factors govern the identity of cell types, which is reflected by genomic enhancer codes. We used deep learning to characterize these enhancer codes and devised three metrics to compare cell types in the telencephalon across amniotes. To this end, we generated single-cell multiome and spatially resolved transcriptomics data of the chicken telencephalon. Enhancer codes of orthologous nonneuronal and γ -aminobutyric acid-mediated (GABAergic) cell types show a high degree of similarity across amniotes, whereas excitatory neurons of the mammalian neocortex and avian pallium exhibit varying degrees of similarity. Enhancer codes of avian mesopallial neurons are most similar to those of mammalian deep-layer neurons. With this study, we present generally applicable deep learning approaches to characterize and compare cell types on the basis of genomic regulatory sequences.

Genomic enhancers form the core of gene regulatory networks (GRNs) that shape and maintain the identity of cell types. GRNs comprise combinations of transcription factors (TFs) that bind to specific transcription factor binding sites (TFBSs) in enhancer regions to regulate the expression of genes. These combinations of TFBSs form enhancer codes that are characteristic for the identity of cell types (1–3). Several tools have been developed to leverage single-cell RNA sequencing (scRNA-seq) and single-cell sequencing assay for transposase-accessible chromatin (scATAC-seq) data to identify cell type-specific GRNs (4, 5). However, characteristic combinations of TFBSs for a cell type are difficult to detect using these methods. Sequence-based deep learning (DL) models have shown major advances to delineate which sequence patterns in enhancer regions are important for cell type-specific chromatin accessibility or gene expression (6, 7). They have contributed substantially to identifying TFBSs specific to mammalian interneurons (8, 9), fly brain cell types (10), mouse liver cells (11), and mouse embryonic stem cells (12). Furthermore, DL models have been applied to predict chromatin accessibility across mammalian brain cell types (13, 14), to com-

pare enhancer codes of melanocytes across species (15), and to identify potential enhancer regions linked to the evolution of neocortex expansion and vocal learning (8, 16). Because these models allow us to identify enhancer codes in cell type-specific enhancer regions, we hypothesized that they may shed light on cell type conservation across species.

Vertebrate telencephala pose ideal examples to study the conservation of enhancer codes because they comprise a variety of cell types that are expected to be maintained by either conserved or diversified GRNs across species (17). The telencephalon constitutes one of the major units of the bauplan for brain development that is shared across all vertebrates (18). Despite shared developmental trajectories, telencephala of different vertebrate brains display a markedly different neuroanatomy (18). As an iconic example, the six-layered neocortex is found in mammals but is absent in other vertebrates (17). Different homologies between structures of vertebrate telencephala have been suggested on the basis of their developmental origin and their projection classes or circuitry (19–21). Single-cell sequencing has been used to compare the transcriptome of cells from the mammalian neocortex, the reptilian three-layered cortex, and three telencephalic nuclei in songbirds (22–24). Both nonneuronal and neuronal cortical cell types have overall conserved molecular identities between the human, marmoset, and mouse, despite transcriptomic differences (24). γ -Aminobutyric acid-mediated (GABAergic) neurons were found to be conserved between the reptilian cortex and mammalian neocortex based on scRNA-seq data,

whereas cross-species relationships of excitatory neurons could not be clearly assigned on the basis of transcriptome comparisons (22). However, groups of turtle excitatory neurons resemble either upper-layer or deep-layer neurons of the mammalian neocortex (22). For songbirds, two nuclei related to vocal learning have been suggested to exhibit similarities to distinct mammalian neocortical neurons, but they do not have the same developmental origin as their potential mammalian counterparts (23). To investigate the similarities between mammalian and avian telencephalic cell types, we compared their transcriptomes to establish a baseline and used genomic sequence-to-function DL models to assess whether these similarities are reflected by shared enhancer codes and conserved TF combinations.

Transcriptome similarities between avian and mammalian telencephalic cell types

To compare cell types on a subclass level and dissect enhancer codes in the mammalian telencephalon, we analyzed five mouse brain and four human cortex or brain single-cell datasets (fig. S1 and table S1) (4, 13, 24–30). In agreement with previous findings, transcriptome comparisons using SAMap recapitulate 1:1 homologies between human and mouse cell types in the cerebral cortex (fig. S2) (24, 31). To compare mammalian cell types with those from another amniote lineage, we generated chicken telencephalon single-cell multiome and spatially resolved transcriptomics [spatial enhanced resolution omics-sequencing (Stereo-seq) and Nova spatial transcriptomics (Nova-ST)] datasets containing 23,179 cells and between 27,487 and 79,104 spatial bins (32, 33).

Besides nonneuronal cell types, we identified seven clusters of GABAergic and eight clusters of glutamatergic neurons, GLU-1 to GLU-7 and immature neurons (IMNs) (Fig. 1, A and B). To ensure the robustness of gene expression profiles of the different cell types, we compared two samples from the chicken telencephalon, which exhibit nearly identical expression profiles (Spearman correlation > 0.9; figs. S3 and S4). We detected the interneuron subclass clusters PVALB+, SST+, and LAMP5+ by their cognate markers and vasoactive intestinal peptide (VIP)-like interneurons by known marker genes, though not by VIP itself (23). Cell types of the striatum, D1 and D2 medium spiny neurons (D1/2MSN), and other striatal-like GABAergic neurons express mammalian marker genes. Next, we investigated whether the different single-cell clusters map to specific regions in the chicken telencephalon using *cell2location* (Fig. 1A and figs. S5 and S6) (29). We verified that the mappings to spatially resolved transcriptomics data and spatial gene expression are robust across clustering resolutions and across the two technologies (Nova-ST and Stereo-seq; figs. S7 to S11). Similar to the

¹Laboratory of Computational Biology, VIB Center for AI & Computational Biology, Leuven, Belgium. ²VIB-KU Leuven Center for Brain & Disease Research, Leuven, Belgium.

³Department of Human Genetics, KU Leuven, Leuven, Belgium. ⁴Center for Molecular Biology of Heidelberg University, Heidelberg University, Heidelberg, Germany.

⁵Department of Neurosciences, KU Leuven, Leuven, Belgium.

*Corresponding author. Email: stein.aerts@kuleuven.be

†These authors contributed equally to this work.

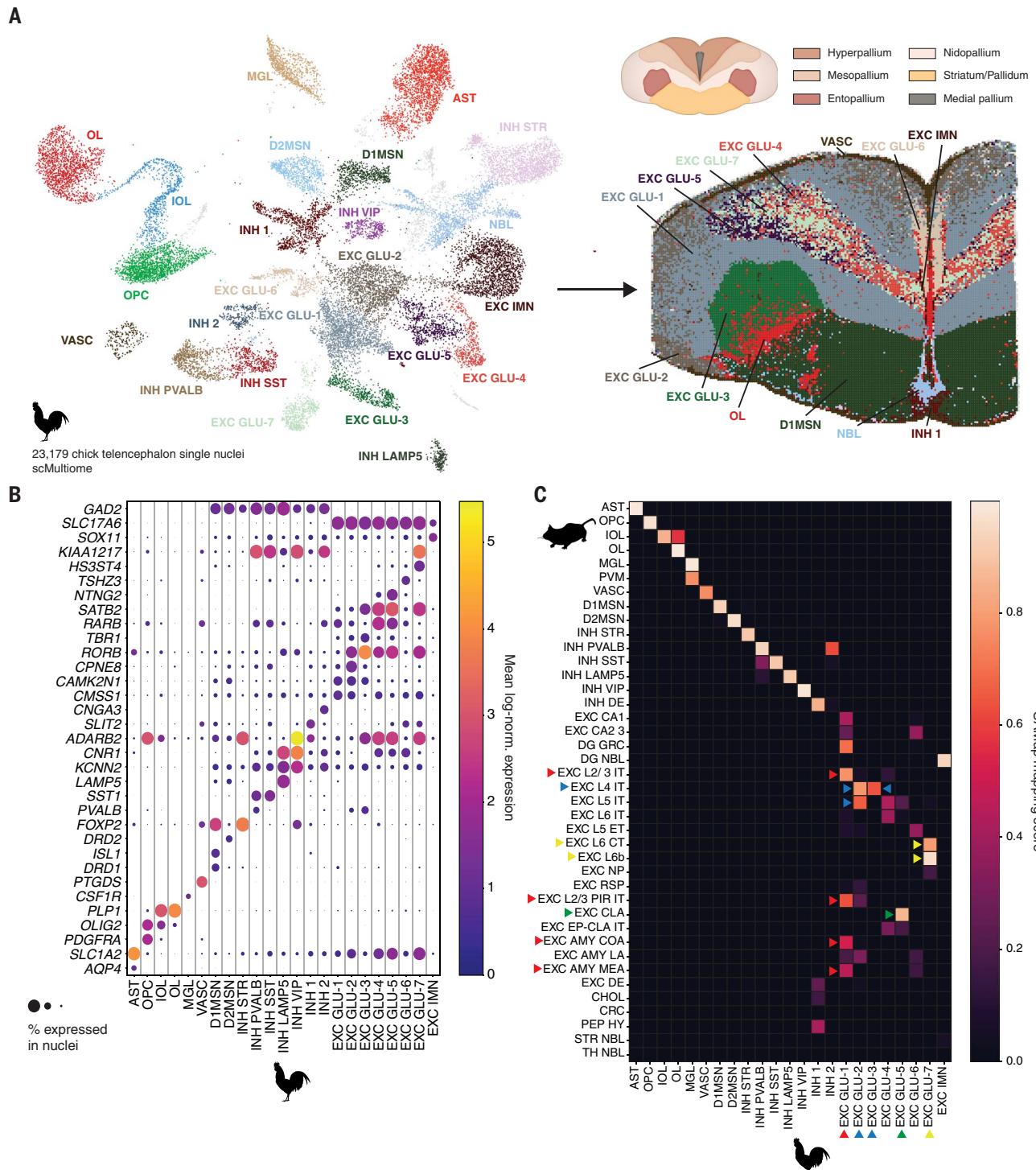


Fig. 1. Cell types of the chicken telencephalon and expression of marker genes. (A) Uniform manifold approximation and projection (UMAP) of single-cell clusters and inferred cell types (left). Single-cell clusters were mapped to spatially resolved transcriptomics data (Stereo-seq) using a bin size approximately equivalent to 50 μm (right). Given that each spatial bin may capture the combined transcriptome of several cells, spatial bins are assigned to single-cell clusters based on the highest density of mapped

clusters labels (right). The pictogram shows anatomical subdivisions of the chicken telencephalon. (B) Expression of characteristic marker genes per chicken telencephalon cell type. Dot sizes indicate the fraction of nuclei of a cluster that express the gene. (C) Transcriptomes of chicken telencephalon cell types were compared with mouse brain cell type transcriptomes using a mouse brain single-nuclei gene expression dataset (29). Triangles highlight relevant similarities for excitatory neurons.

mouse brain, D1/2MSN are exclusively found in the striatum, whereas interneurons preferentially occupy pallial areas (fig. S12A and fig. S13). The eight excitatory neuron clusters exhibit distinct localizations to the nidopallium or hyperpallium (GLU-1 and -2), mesopallium (GLU-4, -5, and -7), entopallium, and medial pallium (GLU-6) (Fig. 1A and fig. S12B), in agreement with proposed subdivisions of the avian pallium (34). The only exception is that we do not observe a split between hyper- and nidopallium given that GLU-1 and GLU-2 are distributed across both the hyper- and nidopallium, in agreement with findings for zebra finches and another recent chicken brain atlas (35, 36). This split is robustly detected across clustering resolutions (fig. S14).

To identify similarities between cell types and to corroborate our marker-based assignments, we used SAMap to match chicken to mouse and human cell types (31). SAMap suggests an unambiguous 1:1 correspondence between mammalian and bird nonneuronal cell types and GABAergic neuron subclasses (Fig. 1C; fig. S2, C to E; and fig. S15). This supports their previously suggested conservation across amniotes (22, 23, 36, 37). Excitatory neuron clusters of the chicken nido- and hyperpallium (GLU-1 and GLU-2) exhibit the highest similarities to mammalian intratelencephalic (IT) neurons of cortical layers L2 and L3 (L2/3 IT) and L4 (L4 IT); excitatory neurons in the L2/3 IT piriiform cortex (PIR); and the cortical, medial, and lateral amygdala (COA, MEA, and LA, respectively), whereas the entopallial RORB+ excitatory neurons (GLU-3) have a moderate similarity to mammalian L4 IT neurons (fig. S12C). Excitatory neurons of the mesopallium (GLU-4, -5, and -7) show the highest resemblances to the mammalian deeper-layer cortical neurons, including L5/6 IT neurons, L6 corticothalamic (CT) neurons, L6b, and neurons of the claustrum (CLA) and CT neurons of the subiculum. Medial pallium neurons (GLU-6) are most similar to mammalian hippocampal cornu ammonis (CA) neurons and the amygdala. IMN and its subclusters exhibit a characteristic expression of *SOX4* and *SOX11*, indicating neurogenesis, and similarities to dentate gyrus (DG) neuroblasts (NBLs) (Fig. 1B and fig. S16) (38). In addition to SAMap, we compared the correlation of 1:1 ortholog expression between the chicken and mouse transcriptomes (fig. S17) (22), confirming the similarities between GLU-1 and mouse L2/3 IT, L2/3 IT PIR and amygdala neurons, and KIAA1217+ GLU-7 and L6CT/L6b neurons. The most prominent transcriptome correspondences are consistently captured across different clustering resolutions and comparisons to mammalian supertypes, which indicates that more coarse-grained cluster resolution sufficiently retains the heterogeneity of the different supertypes [figs. S18 to S21 and supplementary text (39)].

Overall, the transcriptome comparisons find one:many or many:many similarities between excitatory neurons of the chicken and mammalian telencephalon. However, avian meso-, ento-, hyper- and nidopallium, and medial pallium show the highest similarities to either mammalian cortical deep-layer neurons, L4 IT, upper-layer neurons, or neurons of the hippocampus, respectively (Fig. 1C and fig. S2).

Enhancer codes of nonneuronal cell types are highly conserved between birds and mammals

After finding similarities between avian and mammalian cell types based on transcriptomes, the question arises whether these similarities extend to genomic enhancer codes in accessible regions that reflect cell type similarities. Although enhancer regions are not necessarily sequence-conserved across more distantly related species, cell type-specific gene regulation through TFs is often conserved (40–42). If TF-mediated gene regulation is conserved during evolution, we expect selective pressures that preserve specific combinations of TFBSS that form enhancer codes in cis-regulatory regions.

To test this assumption, we used genomic regions with differential chromatin accessibility [differentially accessible regions (DARs)] between telencephalic cell types as a proxy for potential enhancer regions (fig. S22). There is overall low sequence conservation because <25% of the chicken DARs can be lifted over to the mouse or human genome (fig. S23). In line with their assumed earlier evolutionary origin, DARs of nonneuronal cell types and GABAergic neurons show a higher degree of sequence conservation between chicken and mammals compared with pallial excitatory neurons. We verified that the investigated DARs are consistently accessible across different human or mouse datasets and the two chicken telencephalon samples (fig. S24). Because we observed that a low number of cells can have a strong impact on peak calling, we limited the analysis to cell types with at least 350 cells per dataset (fig. S25).

After identifying DARs, we trained DL models to predict cell types directly from DNA sequences of DARs for the different telencephalic cell types (Fig. 2A). To infer robust cross-species predictions, we trained models separately for the two mouse brain (DeepMouseBrain1 and DeepMouseBrain2), two human cortex (DeepHumanCortex1 and DeepHumanCortex2), and the chicken telencephalon (DeepChickenBrain1) datasets (fig. S26). To verify that our models are capable of identifying homologous cell types based on their enhancer codes, we first evaluated whether they recapitulate the cell type homologies between birds and mammals that we would expect based on unambiguous transcriptome comparisons. We evaluated their prediction performance on cell type-specific DARs of nonneuronal cell types and neurons

grouped into broad categories: astrocytes (ASTs), oligodendrocyte precursor cells (OPCs), immature oligodendrocytes (IOLs), oligodendrocytes (OLs), microglia (MGLs), D1/2MSN and striatal-like neurons (MSNs), interneurons (INTs), and excitatory neurons (EXCs). For each chicken cell type, we used the top 100 DARs with the highest log-fold change per cell type and predicted mammalian cell types with the DeepMouseBrain1 and 2 and DeepHumanCortex1 and 2 models (Fig. 2B). Scoring mouse or human regions in the same manner gives identical results (fig. S27, A and B). Similar to the SAMap comparison between chicken and mouse (Fig. 1C), we find near-1:1 correspondences between avian and mammalian nonneuronal cell types directly from accessible DNA sequences. EXCs, MSNs, and interneurons, when grouped together, are also classified correctly. The same regions scored by the human models validate these matches. These findings confirm the transcriptomic cell type matches for nonneuronal cell types and suggest that our models show a robust generalization in classifying previously unseen sequences from different species.

Second, we investigated whether the models that were trained independently on data from different species learned the same sequence features—that is, enhancer codes that correspond to TFBSSs. To identify sequence features, we used both gradient-based contribution scores (43, 44) and in silico mutagenesis (ISM) (45, 46) to derive nucleotide contribution scores at each position of a sequence. These nucleotide contribution scores describe the importance of each nucleotide per sequence for predicting cell type-specific accessibility and allow us to identify important regulatory subsequences, or motifs, corresponding to TFBSSs that are characteristic for cell type identity. As an illustrative example, we investigated the enhancer code of the mouse *fms* intronic regulatory element (FIRE) that regulates the expression of the *Csf1r* gene in MGLs (47). This region has MGL-specific chromatin accessibility, and *Csf1r* is specifically expressed in MGLs (Fig. 2C). The nucleotide importance scores within the FIRE enhancer are highly correlated for all of our models and all species (average pairwise Spearman correlation = 0.45, average $P = 1.78 \times 10^{-11}$) and detect previously validated TFBSSs as the most important features of the sequence (Fig. 2D). Nucleotide importance scores derived from other models with different architectures and training methods—namely Enformer and a ChromBPNet model that we trained on the MGL ATAC profile—agree with the ones obtained from our models (Fig. 2D) (48, 49). To further validate the model predictions, we performed in vitro saturation mutagenesis using massive parallel reporter assay (MPRA) in the mouse BV2 cell line. The in vitro enhancer activity profile of this MGL cell line shows decreases in activity

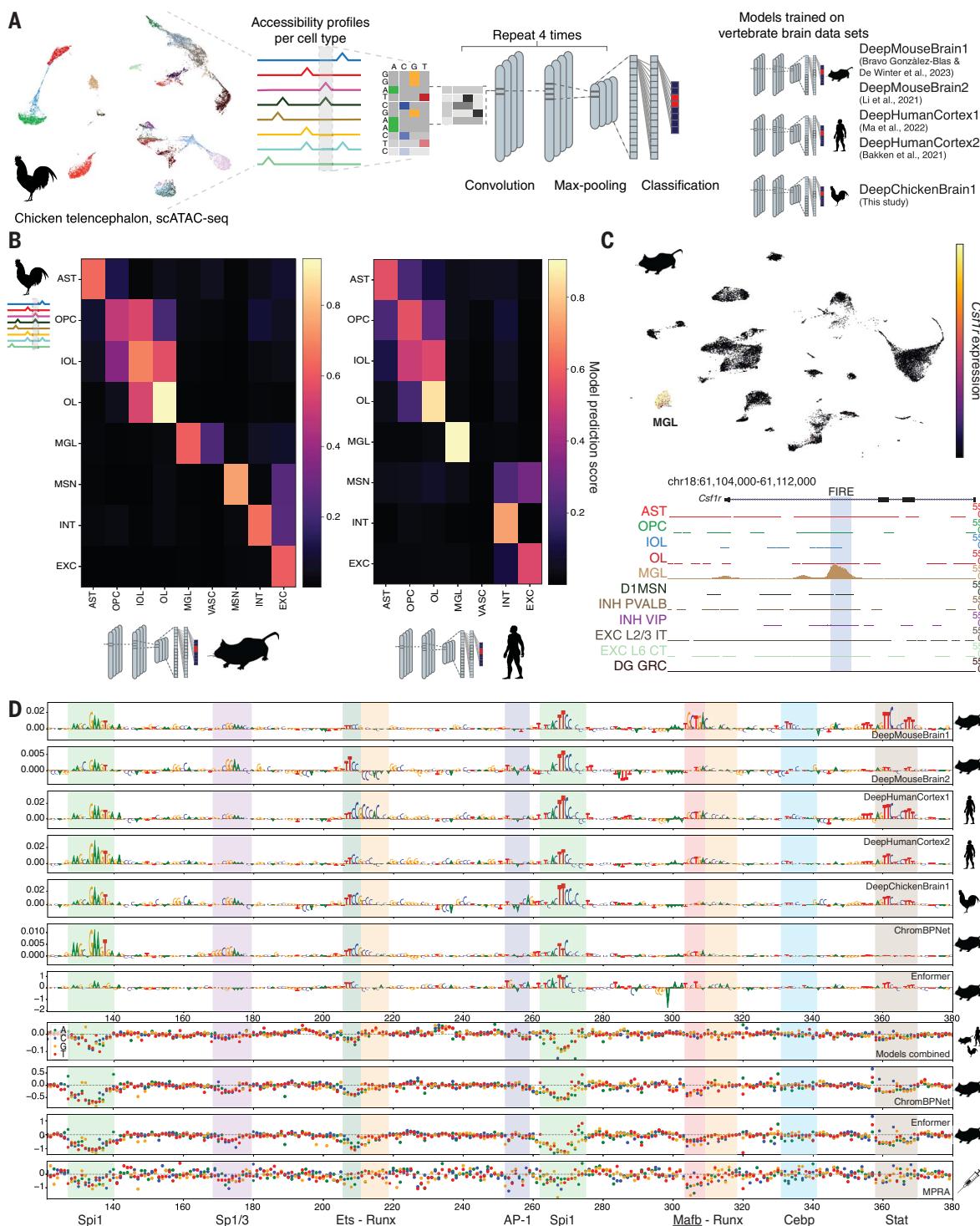


Fig. 2. Cross-species enhancer modeling. (A) Overview of enhancer model architecture and workflow. For each of the five brain datasets, a model was trained. (B) Cross-species model predictions from the mouse (left) and human (right) models on the top 100 DARs per chicken cell type. The median prediction score per region is shown. Both mouse and human predictions are the consensus of the two respective models per species. EXC, MSN, and INT cell types are grouped into broad cell type categories. (C) scRNA-seq UMAP of mouse brain cells from the DeepMouseBrain1 dataset (4) showing the MGL-specific expression of the *Csf1r* gene (top). The cell type-specific chromatin accessibility of the

MGL enhancer FIRE (mm10 chr18:61108475-61108975), regulating *Csf1r* expression in MGLs, is depicted underneath. (D) Nucleotide contribution scores of FIRE obtained from the five enhancer models (order: DeepMouseBrain1, DeepMouseBrain2, DeepHumanCortex1, DeepHumanCortex2, DeepChickenBrain1), from a MGL ChromBPNet model, and from Enformer are shown. ISM scores are shown for a consensus of all five enhancer models, the ChromBPNet model, and Enformer. Finally, in vitro mutagenesis from MPRA activity in mouse BV2 cells is shown. Previously identified TFBSSs are highlighted, as well as an uncharacterized potential Mafb binding site.

specifically for nucleotide changes that disrupt predicted TFBSs and is correlated with the predicted importance of the nucleotides (Spearman correlation with the composite ISM = 0.26, $P = 1.04 \times 10^{-17}$) (fig. S28). This confirms that for this enhancer, the nucleotide importance scores, which our models learned to be characteristic for MGL enhancers, are affecting the *in vitro* enhancer activity. They further support that the previously experimentally identified TFBSs in this enhancer are functional (Fig. 2D) (47). Additionally, we find a previously unidentified MAFB-like TFBS in all mammalian models that is confirmed by the MPRA activity. These results indicate that our models independently learned the same features for explaining their predictions for the mouse MGL enhancer. This suggests conservation of the MGL enhancer code across the three amniote species and provides confidence in our models to enable cross-species comparisons and to detect important cis-regulatory features.

Given that our models trained on data from different species and learned similar enhancer codes for multiple cell types, we reasoned that we could use them for assessing homologies between cell types. We illustrate this for a candidate enhancer in mouse ASTs. This region is located inside an intron of *Prdm16* and is conserved (sequence identity 60%) and specifically accessible in both chicken and mouse ASTs (log-fold change 5.33 and 5.32, respectively) (Fig. 3A). During development, *Prdm16* is expressed in radial glia and contributes to cell migration through transcriptional silencing (50). In adult mammalian and avian brains, *Prdm16* is a characteristic AST marker gene (Fig. 3B). All models accurately predict this region as being specific to ASTs (Fig. 3C and fig. S29A). To compare the similarity of learned AST enhancer codes between our chicken, human, and mouse models, we computed nucleotide contribution scores for all cell types and compared them with contribution scores derived from the other models. This allowed us to assess which cell types from two different species yield the most similar enhancer codes. For example, when comparing nucleotide contribution scores for this region derived from DeepMouseBrain1 for AST with the scores of all cell types derived from DeepChickenBrain1, the chicken AST class has the highest degree of similarity to the mouse AST class (Spearman correlation coefficient = 0.63, $P = 9.81 \times 10^{-55}$). Similar to the MGL enhancer (Fig. 2D), the sequence features detected with high importance scores by all models correspond to potential TFBSs (Fig. 3C and fig. S29B). In addition, sequence alignment between the mouse and chicken regions shows that point mutations and insertions occur at nucleotides with low importance scores, maintaining TFBS identity. The same applies for the importance scores of the chicken homologous region (fig.

S29C). As a negative control, contribution scores for the chicken IOL cell type do not indicate any important TFBSs in this region and are anticorrelated with the mouse AST contribution scores (Spearman correlation = -0.25, $P = 1.66 \times 10^{-8}$) (Fig. 3C). Hence, the correlation between nucleotide contribution scores can be used to compare for which cell types two different models learned the most similar enhancer codes, which provides an indication about cell type homologies. Next, we applied this metric of nucleotide importance score correlation to compare enhancer codes pairwise across all mammalian and avian telencephalic cell types. The correlation between contribution scores is higher for homologous cell types than for nonhomologous cell types, including when the comparison is done on the top mouse and human DARs (Fig. 3D and fig. S27, C to E). This strategy enables cell type comparison based on cell type-specific nucleotide contribution scores, which represent the underlying gene regulatory code that the models learn.

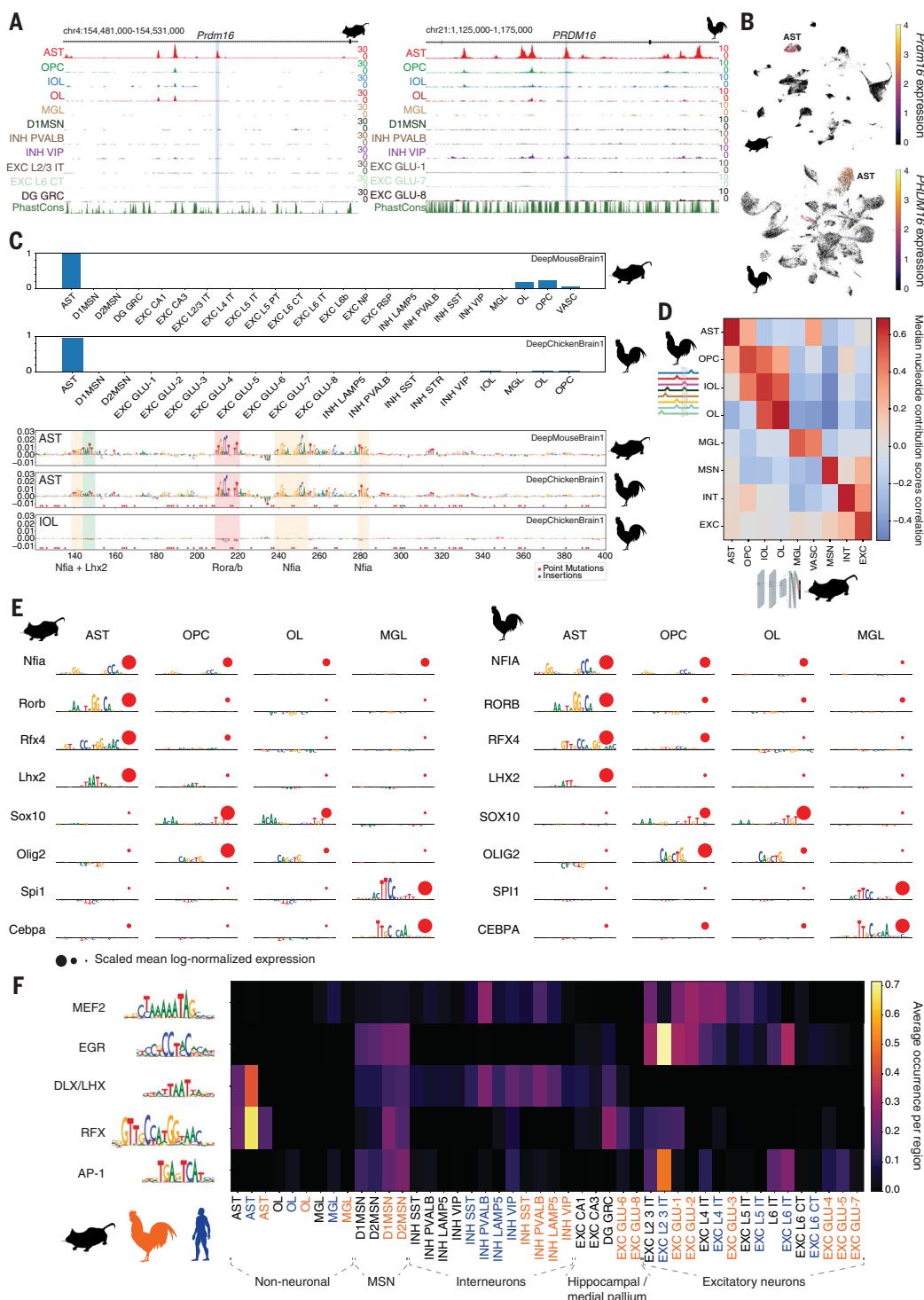
As an alternative strategy to assess cell type similarity, we compared the learned TFBSs per cell type. We used TF-MoDISco (51) to infer cell type-characteristic sequence patterns for each set of cell type-specific DARs (log-fold change = 1.5; table S2). To identify learned patterns that are shared across cell types and species, we compared the learned patterns of all models against each other using the MEME suite and clustered them into 42 groups of motifs (fig. S30) (52, 53). Additionally, we screened all learned patterns against the *cistarget* database (4). The majority of learned patterns resemble TFBSs (54). The learned TFBS patterns correlate with differential TF expression across ASTs, OLs, and MGLs (Fig. 3E). For example, concordant patterns and expression of genes are found for NFIA, RORB, RFX4, and LHX2 in ASTs; SOX10 and OLIG2 in OLs; and Sp1 and CEBPA in MGLs. In agreement with AST-specific TFBSs that we identified with TF-MoDISco, the potential AST enhancer region within the *PRDM16* intron harbors NFIA, RORB, and LHX2 TFBSs that our models consider important for identifying this region as AST specific (Fig. 3A). To evaluate whether the learned TFBS motifs can be used to distinguish different cell types, we evaluated the correlation of their average number of instances across cell types and species (fig. S31). We then performed hierarchical clustering based on the correlation of motifs between the human, mouse, and chicken cell types. The cell types of the three different species cluster into ASTs, OLs, MGLs, interneurons, D1/2MSN, and excitatory neurons based on the correlation of learned TFBS motifs, which indicates shared enhancer codes across the species (fig. S32). These broad cell type categories show characteristic preferences for TFBS motifs of MEF2, EGR, DLX/LHX, RFX, and AP-1 family TFs (Fig.

3F). The comparison of patterns thus represents a third strategy to use enhancer models across species, besides the model prediction scores and the correlation of attribution scores.

Enhancer codes of GABAergic neurons are conserved between birds and mammals

Using the same three strategies from the enhancer models, we investigated avian and mammalian D1/2MSN and interneurons in more detail and assessed whether we can recapitulate the conservation of GABAergic neurons at the level of enhancer codes. The five previously mentioned scATAC-seq datasets were limited in their resolution to distinguish highly similar cell types, such as interneuron subclasses (i.e., SST+ from PVALB+ and LAMP5+ from VIP+) (fig. S33), which reduces the ability of the models trained on those datasets to distinguish them based on sequence (figs. S34 and S35). To overcome this resolution problem, we trained two new models using the CREsted package (55), with a different architecture and training procedure, on higher-resolution, large-scale atlases of full mouse (DeepMouseBrain3) (13) and human brain (DeepHumanBrain) (28) scATAC-seq datasets (figs. S36 and S37 and table S3). Additionally, we trained a chicken brain CREsted model (DeepChickenBrain2) on the chicken scATAC-seq dataset. An overview of the different models can be found in tables S4 to S12. The DeepMouseBrain3 and DeepHumanBrain models are able to differentiate D1/2MSN from striatal-like inhibitory neurons (INH STR) and cell types within interneuron classes on the basis of predictions and nucleotide contribution score correlation (Fig. 4A and fig. S38, A to C). To further assess the issue of limited scATAC resolution in interneurons, we calculated highly specific DARs from a dedicated mouse interneuron dataset (56) and show that these models can better differentiate interneurons (fig. S38, D to F, and fig. S39).

Learned enhancer codes for predicting D1/2MSN and interneurons correspond to binding sites for key TFs and are correlated with TF expression (Fig. 4B, fig. S40, and table S13). Particularly, MEIS1, EGR3, PBX2/3, and HLF TFBSs are characteristic for spiny neuron enhancers. By contrast, basic helix-loop-helix (bHLH) factor TFBSs, likely corresponding to TCF4 and TFBS of NFIA/B/X, are characteristic for interneurons. For interneurons, potential NFIA/B/X TFBSs are more specific to LAMP5+ and VIP+ interneurons, whereas potential MAFG TFBSs are more characteristic for the PVALB+ and SST+ interneurons. Homeo-domain factor TFBS, which could correspond to different DLX-family TFs (DLX6 in Fig. 4B), and AP-1 TFBS, such as potential JUN or FOS family TFBSs, are characteristic for all D1/2MSN and interneurons. DLX family TFBSs are characteristic for cortical GABAergic neurons

**Fig. 3. Conservation of enhancer code in vertebrate brain cell types.**

(A) scATAC tracks of the *PRDM16* region in chicken (galGal6 chr21:1148446-1148946) and mouse (mm10 chr4:154506677-154507177). (B) UMAPs showing *Prdm16* expression in mouse and chicken telencephalon. (C) DeepMouseBrain1 and DeepChickenBrain1 predictions and contribution scores for the mouse region. Point mutations and insertions are indicated in red and blue. High contribution score TFBSs and potential TFs are indicated based on their similarity to known TFBS motifs and cell type–specific expression. (D) Consensus median cross-species nucleotide contribution scores Spearman correlation between DeepMouseBrain1 and 2 and DeepChickenBrain1 for the top 100 DARs per

chicken cell type. EXCs, MSNs, and INTs are grouped into broad cell type categories. (E) Contribution scores of characteristic sequence patterns for mouse (left) and chicken (right) f cell types. Potential TFs are indicated based on known TFBSs and the correlation of their expression with importance scores. The scaled mean TF expression per cell type is shown by red circles. (F) Heatmap depicting the average number of instances of sequence patterns in DARs that are characteristic for broad cell type categories for DeepMouseBrain1 and 2, DeepHumanCortex1 and 2, and DeepChickenBrain1. Cell types of the different species are indicated in black, orange, and blue for mouse, chicken, and human, respectively.

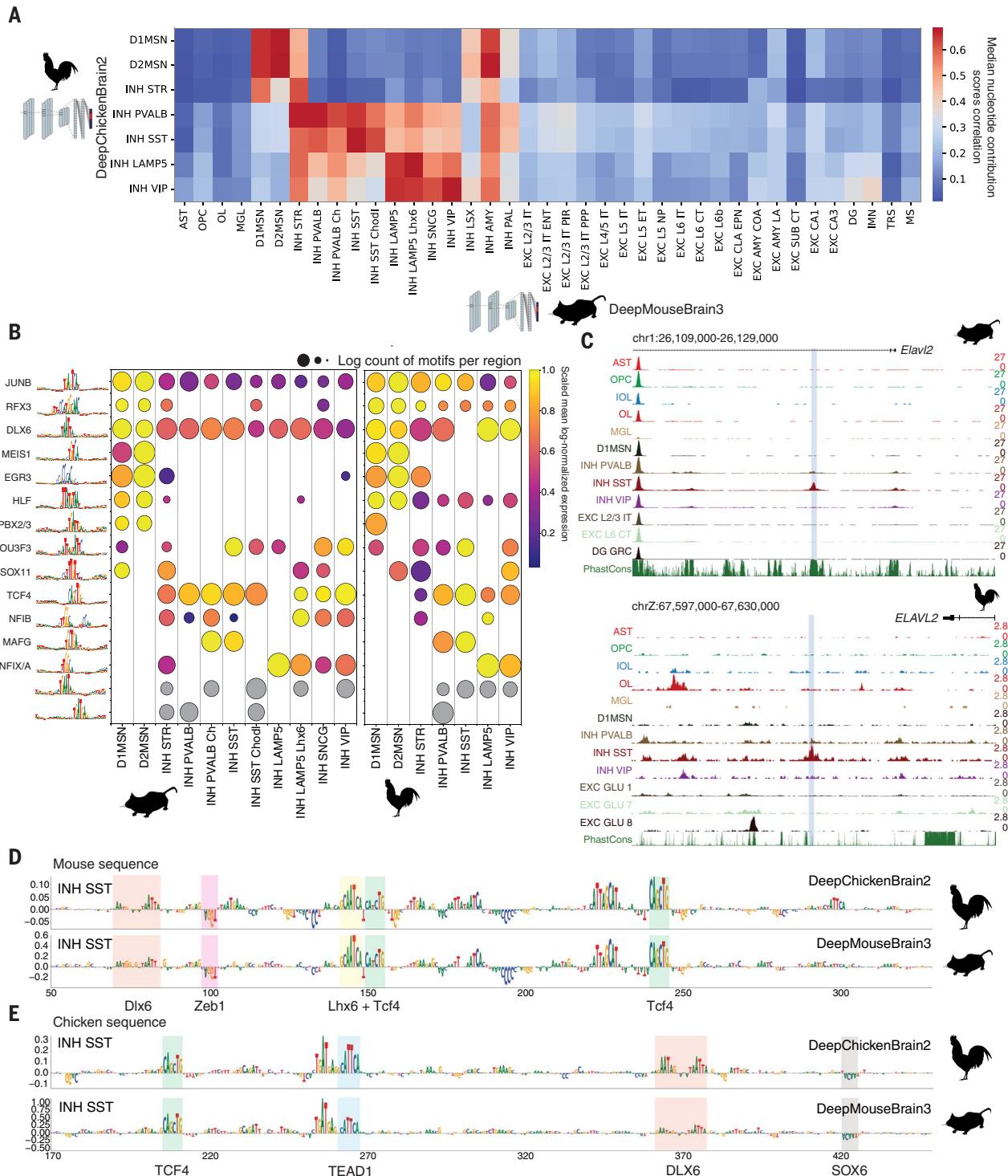


Fig. 4. Conserved enhancer codes of GABAergic neurons across mouse and chicken. **(A)** Spearman correlation between the nucleotide contribution scores of the DeepMouseBrain3 and DeepChickenBrain2 models for the top 100 DARs of chicken medium spiny neurons (MSNs) and interneurons. **(B)** Importance of shared TFBSS that were learned by DeepMouseBrain3 and DeepChickenBrain2 based on their occurrence in the top 1000 DARs per cell type for MSNs and interneurons for mouse (left) and chicken (right). The sizes of the dots show the \log_2 of their occurrence. On the left, potential TFs whose expression is correlated with the occurrence of the TFBSS are

indicated based on known TFBS motifs. The colors of the dots show the scaled mean expression (26). Gray circles indicate that no fitting TF match was found. **(C)** scATAC tracks of the chicken (bottom) and mouse (top) *ELAVL2* enhancer candidates (galGal6 chrZ:67613064-67613564 and mm10 chr4:91392313-91392813) showing somatostatin (SST) specificity. **(D)** DeepMouseBrain3 and DeepChickenBrain2 contribution scores for their SST predictions on the mouse *Elavl2* region. **(E)** DeepMouseBrain3 and DeepChickenBrain2 contribution scores for their SST predictions on the chicken *ELAVL2* region.

(3). To illustrate how these TFBSSs are arranged in potential interneuron enhancer regions, we investigated regions adjacent to the interneuron marker gene *ELAVL2* that are specifically accessible in SST+ interneurons in both chicken and mouse (Fig. 4C). For these instances,

our models consistently predict TFBS motifs that we identified to be part of the conserved enhancer codes of SST+ interneurons (Fig. 4, B, D, and E, and fig. S4I). As an example for a D1/2MSN-specific putative enhancer, we identified a region in an intron of *FOXP2* that is

specifically accessible in D1/2MSN and INH STR. This region is highly conserved in sequence (sequence identity 94.9%) and ATAC signal (log-fold change 3.66 and 3.80) between mouse and chicken (Fig. 5, A and B). *FOXP2* is a marker gene for D1/2MSN (Fig. 5A) and

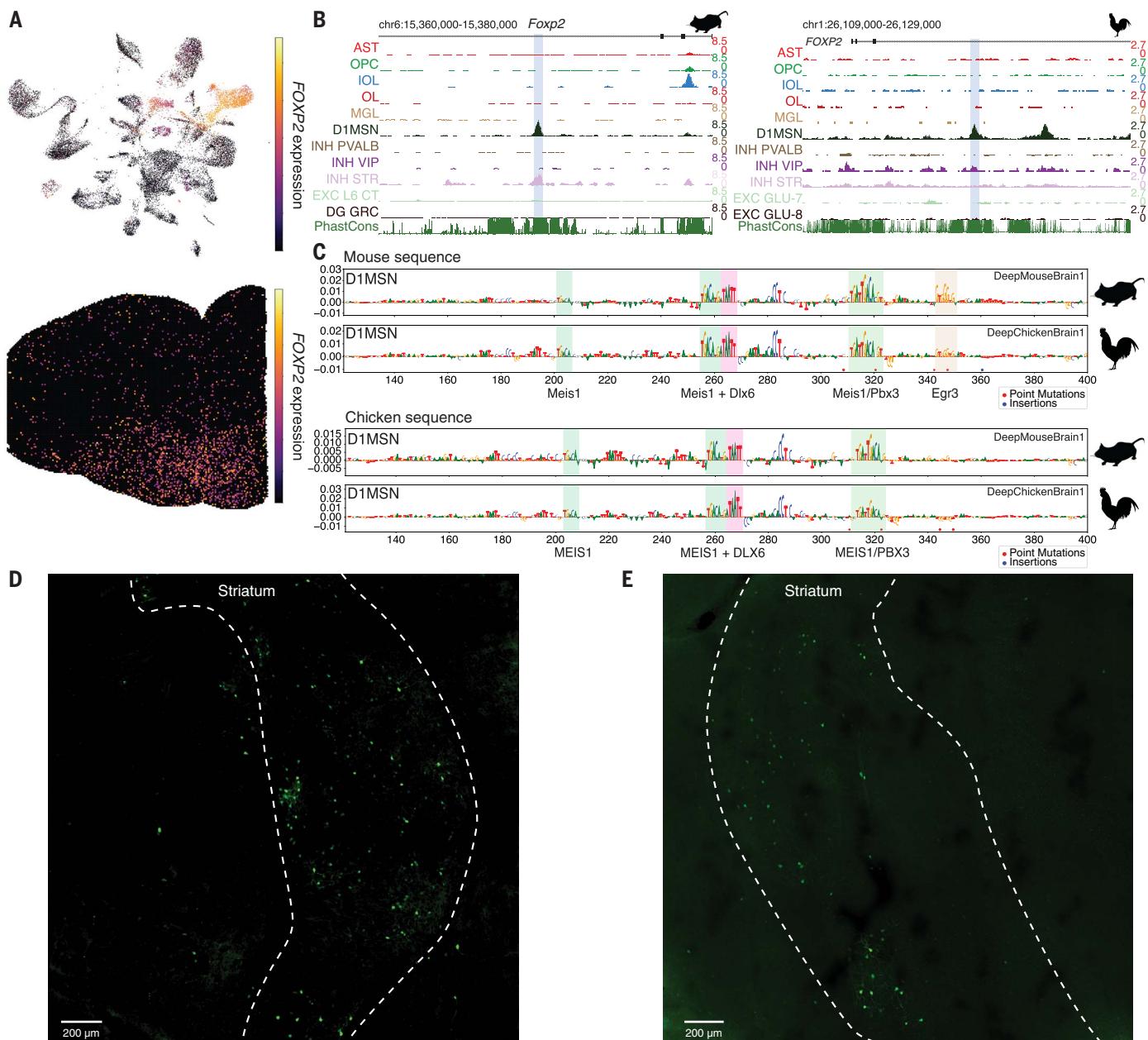


Fig. 5. Identification of an intronic medium spiny neuron enhancer inside *Foxp2*. **(A)** scRNA-seq UMAP showing that *FOXP2* expression is specific to D1MSN and striatal-like inhibitory neurons (INH STR) in chicken telencephalon single-cell data (top) and its highest expression in striatum in spatially resolved transcriptomics (Stereo-seq) data (bottom). **(B)** scATAC tracks of the *FOXP2* region in chicken (galGal6 chr1:26118953-26119453; left) and mouse (mm10 chr6:15369301-15369801; right) showing D1MSN specificity. **(C)** DeepMouseBrain1 and DeepChickenBrain1 contribution scores for their D1MSN classes for both the mouse (top) and chicken sequence (bottom); shown is the reverse complement of the mouse sequence. Point mutations and

insertions in the mouse sequence when aligned to the chicken sequence are indicated in red and blue, respectively, and vice versa in the chicken sequence. In the chicken, the Egr3 binding site present in the mouse is destroyed. TFs that potentially bind to nucleotide features with high contribution scores are indicated and assigned on the basis of their similarity to known TFBS motifs and cell type-specific expression. (**D** and **E**) GFP expression in the mouse brain driven by the D1/2MSN enhancer sequence within the mouse *Foxp2* intron (D) and the orthologous sequence within the chicken FOXP2 intron (E). The striatum is delimited by dashed lines. Slight expression is also observed in the thalamus.

is important for plasticity in the striatum, is important for striatum-dependent learning of skills including speech, and is linked to the evolution of language in humans (57–59). The mouse, human, and chicken models detect MEIS1 and homeodomain TFBSSs to be characteristic of D1/2MSN, in agreement with the patterns identified by TF-MoDISco (Fig. 3F). We assessed the activity of this chicken enhancer candidate *in vivo* by incorporating it into an adeno-associated virus (AAV) vector along with a green fluorescent protein (GFP) reporter and administering it to a mouse brain by means of injection. The reporter specifically expressed GFP in the striatum and striatum-like amygdalar nucleus, in agreement with our predictions and observed scATAC signal (Fig. 5D; fig. S42, A and B; and fig. S43). Our analyses suggest that enhancer codes of GABAergic neurons are conserved between mammals and birds.

Enhancer codes suggest similarities between mammalian and avian excitatory neurons

In contrast to the conservation of GABAergic neurons, homologies between avian pallial and mammalian cortical excitatory neurons are highly debated (60). Building on our findings for nonneuronal and GABAergic neurons, we investigated whether the enhancer models confirm the similarities between excitatory neuron transcriptomes and may provide additional insights into their conservation or divergence.

To this end, for the top 100 DARs per excitatory neuron type in chicken, we calculated cross-species predictions and nucleotide contribution score correlations with mouse and human cell types present in the DeepMouseBrain, DeepHumanCortex, and DeepHumanBrain models (Fig. 6A and figs. S34, S35, S44, and S45). We also assessed the influence of the 100-DARs threshold on cross-species comparisons and found that increasing it to 500 recapitulated our findings (fig. S46). Additionally, we verified that forced subclustering of the chicken GLU subclasses [see supplementary text (39)] could not identify more specific matches (fig. S47). As was observed for excitatory neurons as a broad category (Fig. 3D), the enhancer code of chicken excitatory neuron clusters only correlates with mouse excitatory neurons. Similar to the transcriptome comparisons (Fig. 1C), the DL models suggest one:many or many:many correspondences between the enhancer codes of most avian and mammalian excitatory neuron subclasses. The strongest correspondences of enhancer codes are concordant with transcriptome matches, namely between mammalian upper-layer neurons, neurons of the amygdala, entorhinal cortex (ENT), and L2/3 IT PIR neurons with avian hyper- or nidopallium (GLU-1) and between L6 CT and L6b neurons with avian mesopallium neurons (GLU-7). The GLU-3, -5, and -6 matches are

also concordant with the transcriptomic comparison. Conversely, the GLU-1 match to LA neurons by enhancer codes is not recapitulated by the transcriptome, which finds a stronger match to amygdalar COA neurons. In contrast with findings between mammalian neurons (61), the predicted correspondences do not clearly coincide with mammalian projection classes or localization to cortical layers (fig. S48).

To investigate how these correspondences are reflected by learned TFBSSs, we identified TFBSS patterns across species that are important for differentiating between excitatory neuron subclasses, in the same manner as we did for GABAergic and nonneuronal cell types above. The most cell type-characteristic learned motifs resemble binding sites of RFX3, RORB, TBR1, POU3F1, LHX2, and NR4A2 and bHLH motifs. bHLH binding sites that likely correspond to the TCF-family, ATOH-family, or NEUROD-family TFs are shared by most excitatory neurons (Fig. 6B and fig. S49). Potential binding sites of RFX3 are most characteristic for mammalian neocortex and PIR L2/3 IT neurons, neurons in the amygdala, avian hyper- or nidopallium neurons (GLU-1), mammalian neurons of the hippocampus or DG, and avian medial pallium (GLU-6), in agreement with their correlated enhancer codes (Fig. 6A). The learned importance of RORB and TBR1 motifs is correlated with their expression across cell types (fig. S9, C and D). Potential binding sites of TBR1 are learned to be most characteristic for L6 CT and b, L2/3 IT PIR neurons, and CLA neurons in mouse regions. For chicken regions, DeepChickenBrain2 learned potential TBR1 binding sites to be characteristic for the entopallial GLU-3 and nidopallial GLU-2 neurons. Both the mouse and chicken models learned bHLH TFBSSs likely corresponding to NEUROD1/2 to be important for most cell types, with highest occurrence in the hippocampus and DG and medial pallium neurons. Potential binding sites of NFIA/B are most characteristic for mammalian L6 CT and avian mesopallium neurons. Our models successfully identified many TFBSSs of activator or chromatin-opening TFs that are characteristic for mammalian neurons. Therefore, we investigated the influence of known repressors or potential chromatin-closing TFs, such as CUX1, SATB2, FEZF2, and BCL11B (CTIP2 in human) (62). Although our models do not suggest that SATB2, FEZF2, or BCL11B binding influence cell type-specific chromatin accessibility, CUX1 patterns were detected for L4/5 IT in mouse and L2/3 IT in human. *In silico* motif insertion experiments confirmed CUX1 to be important for L2/3 IT in mouse as well (fig. S50) (54).

To validate the enhancer code-derived correspondences, we performed *in vivo* enhancer reporter assays in the mouse brain for two predicted chicken enhancer regions: a region

specifically accessible in GLU-7, predicted to be similar to mammalian L6CT/L6b neurons, and a GLU-1 region predicted to resemble L2/3 IT, PIR, and amygdala neurons in mammals (Fig. 6, C and D). For GLU-7, we tested a predicted enhancer upstream of the GLU-7 marker gene *KIAA1217* (fig. S51, A and B), which shows specific activity in the deep layer of the mouse neocortex (Fig. 6E). Most GFP-positive cells are *Bcl11b* positive (CTIP2 antibody), which confirms their deep-layer excitatory neuron identity (Fig. 6F and fig. S53A) (62). The enhancer also shows strong activity in the hippocampus, which was predicted by DeepMouseBrain3 (fig. S51B), and in the thalamus (fig. S42C). The hippocampal activity pattern has been shown to be characteristic to AAV-PHP.eB infections and may not reflect cell type-specific activity in the hippocampus (63, 64). Moreover, we identified a mouse enhancer candidate near the ortholog of *KIAA1217*, *Etl4*, indicating conservation of enhancer code independently of sequence conservation (fig. S51, C to E). For GLU-1, we tested the activity of a predicted enhancer near *ZNF804B* (fig. S52A), which drives GFP expression in neocortical L2/3 IT, L2/3 IT piriform cortex, and sparsely in the amygdala (Fig. 6G and fig. S42D), as predicted by the enhancer models (fig. S52, B and C) and transcriptome-based comparisons (Fig. 1C). By staining against SATB2 (62), we confirmed that the GFP-positive cells are neocortical excitatory neurons (Fig. 6H and fig. S53B). Chicken and mouse models detect cell type-specific TFBSSs in these enhancers that highlight the conservation of the enhancer codes (Fig. 6, B, I, and J).

Overall, the analysis of enhancer codes and transcriptome comparisons suggest different degrees of similarity between avian and mammalian cell types. We present four approaches for computing cell type similarities: (i) transcriptome comparison, (ii) predictions from sequence-based models, (iii) correlation of derived nucleotide contribution scores, and (iv) similarities of TFBSS motifs. To summarize the similarity between avian and mammalian cell types, we aggregated these four cell type similarity metrics into an average cell type similarity (Fig. 7, A and B, and figs. S54 and S55). The combined similarity confirms 1:1 correspondences of nonneuronal cell types, D1/2MSN, and interneurons. The same strategy applied to human-mouse comparison suggests near-1:1 correspondence between human and mouse excitatory neurons, corroborating the enhancer metric validity (Fig. 7C). Between mammal and chicken excitatory neurons, the highest similarities are found between GLU-7 and L6 CT and b and between GLU-1 and L2/3 IT, amygdala, PIR L2/3IT, and ENT neurons.

Discussion

We trained independent DL models on mammalian and avian genomic sequences to identify

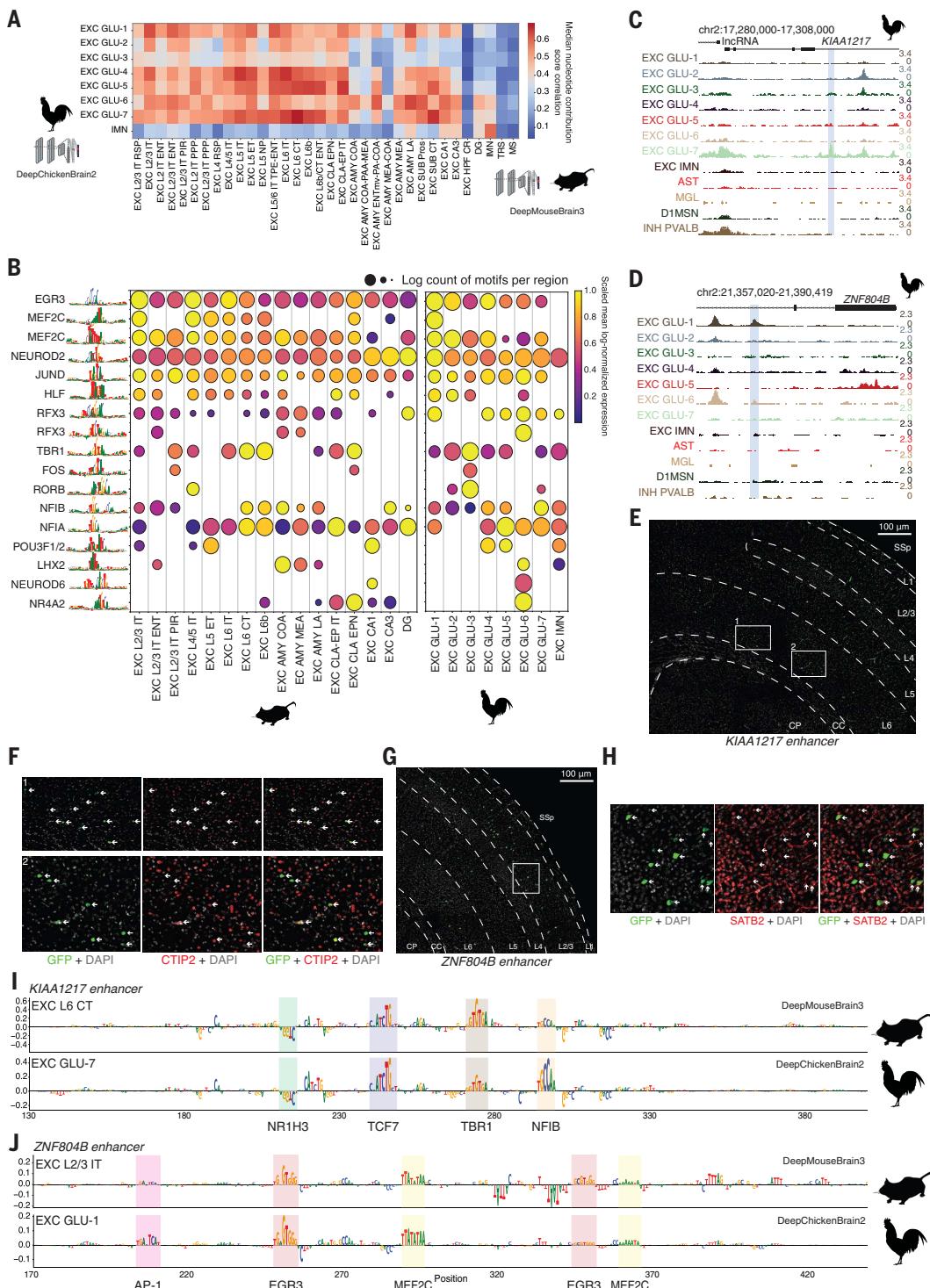


Fig. 6. Enhancer code of excitatory neurons in the chicken and mouse.

(A) Heatmap of the median nucleotide contribution score Spearman correlation between DeepChickenBrain2 and DeepMouseBrain3 for the top 100 chicken EXC DARs. (B) Importance of TFBSs for DeepMouseBrain3 and DeepChickenBrain2 based on their occurrence in the top 1000 DARs. The dot sizes indicate the \log_2 of their occurrence. Potential TFs are indicated based on known TFBSs and their expression. The color gradient shows the scaled mean expression of TFs. (C to E) scATAC profiles of the KIAA1217 locus (C) showing specific accessibility in GLU-7 (D) and of the ZNF804B locus in GLU-1 (E). GFP expression is driven by

the KIAA1217 enhancer in the mouse primary somatosensory cortex (SSp), the corpus callosum (CC), and the caudoputamen (CP). (F) CTIP2 staining and GFP expression driven by the GLU-7 KIAA1217 enhancer. Eleven out of 18 GFP-positive cells (arrows) show CTIP2 expression. DAPI, 4',6-diamidino-2-phenylindole. (G) GFP expression driven by the ZNF804B enhancer in the mouse SSp. (H) SATB2 staining and GFP expression driven by the ZNF804B enhancer. Nine out of the 12 GFP-positive cells (arrows) show SATB2 expression. (I and J) Nucleotide contribution scores for the KIAA1217 enhancer candidate for L6 CT and GLU-7 (I) and of the ZNF804B enhancer candidate for L2/3 IT and GLU-1 (J).

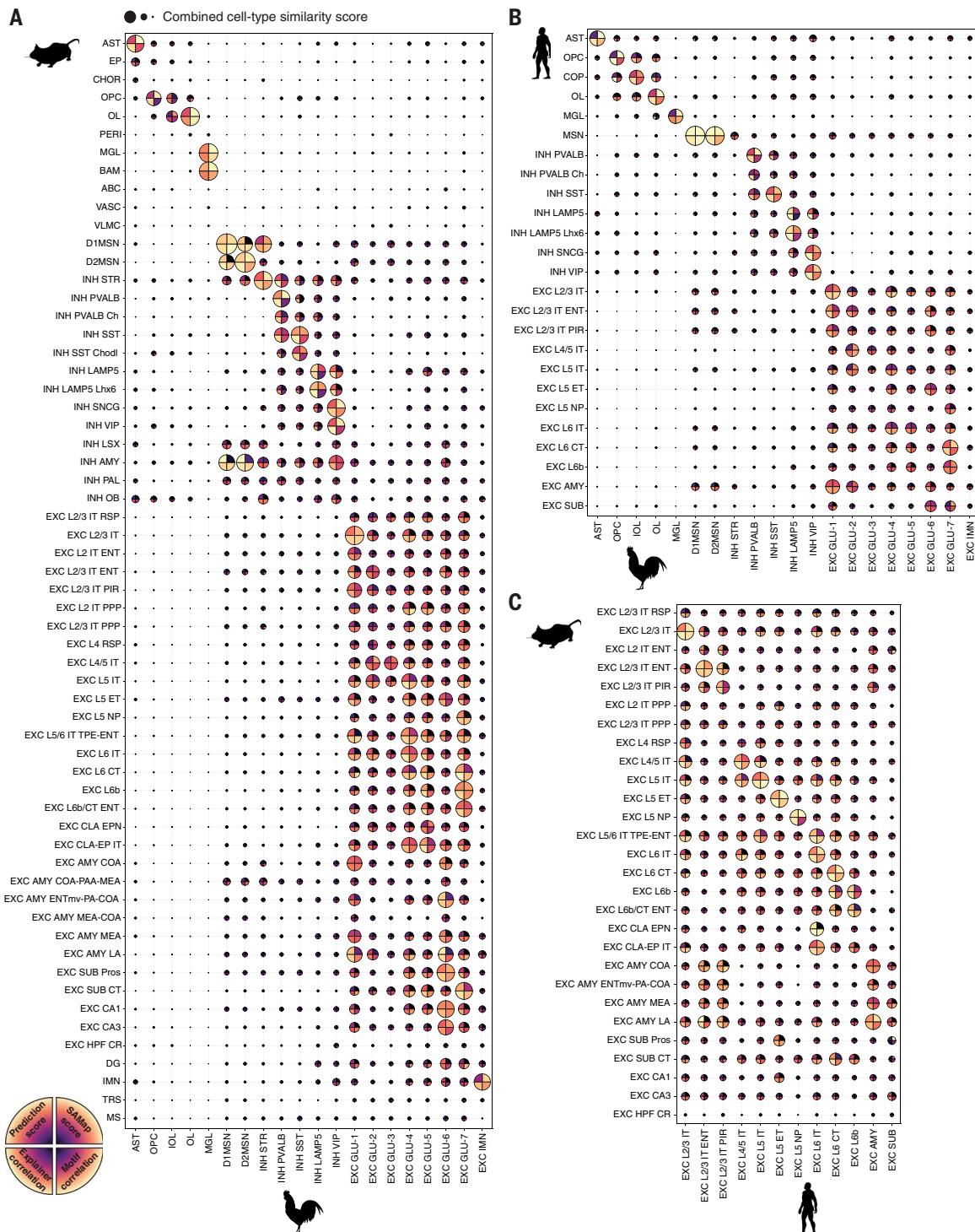


Fig. 7. Overview of regulatory code-matching methods between chicken-mouse and chicken-human brain cell types and mouse-human glutamatergic cell types. (A to C) Combined cell type similarity score consisting of the SAMap score, prediction score, contribution score correlation, and motif correlation between chicken and mouse (A), chicken and human (B), and human and mouse (C) cell types. The circle sizes scale with the aggregate of the four metrics. SAMap scores are obtained from the mappings between the Yao *et al.* mouse brain (26), the Siletti *et al.* human brain (27), and the chicken brain dataset. DL predictions,

contribution score correlations, and motif correlations are derived from DeepChickenBrain2, DeepMouseBrain3, and DeepHumanBrain. The predictions are the reciprocal average median of scores on chicken DARs and mouse DARs (A), chicken DARs and human DARs (B), and mouse DARs and human DARs (C) based on the Zu *et al.* mouse brain (13), the Li *et al.* human brain (28), and the chicken telencephalon datasets. Contribution score correlations were obtained from chicken DARs in (A) and (B) and mouse DARs in (C). All scores were standardized between 0 and 1. Negative correlation scores were set to zero.

enhancer codes that are sufficient to characterize major brain cell types. These enhancer codes comprise TFBS combinations and arrangements that are characteristic for the identity of the different cell types in the telencephalon and are conserved across mammals and birds. Our results provide the cis-regulatory complement to previous findings that TFs are often found as main components of conserved core regulatory circuits that define animal body plans and cell type identity (40–42). Although gene expression and morphological properties of homologous cell types may show larger variation between species, core sets of TFs have been found to show a high degree of conservation. For instance, developmental programs of cells in the retina are regulated by a set of key TFs that is conserved across vertebrates (65, 66). Also, the key TFs underlying heart development are conserved from the fruit fly to the human (67). Genomic signatures of such TF combinations that form enhancer codes can be conserved across evolutionary distances as far as fishes and sponges (68). Hence, enhancer codes associated with these TF combinations can be used as a tool for assessing cell type homology. The conservation of enhancer codes is the foundation for training and cross-species scoring of cell type-specific models on accessible chromatin regions. Such models have been successfully applied to predict chromatin accessibility for primate and mouse neocortical cell types, which indicates the conservation of the mammalian neocortical enhancer code (13, 14).

We present three computational strategies on how to use the conservation of enhancer codes to assess cell type similarities between species, as a complement to transcriptome-based similarities. These strategies can reveal conserved enhancer codes without requiring sequence or target gene conservation. Consequently, this approach can be used for comparisons between evolutionarily distant species, such as mammals and birds, even if very few or none of the enhancer regions are alignable by genomic sequence. Although the applicability of scATAC-seq-based modeling may be limited by the number of cells per cell type and the resolution of clusters, there is an overall agreement with transcriptome-based comparisons (fig. S55). Furthermore, we found that enhancer code-based metrics perform better on capturing global similarities between cell types across species (fig. S56). Overall, nucleotide contribution score correlation is the most robust enhancer code-based metric. Advances in scATAC-seq technologies will resolve the limitations of comparing rare cell types and states in future datasets because the number of sequenced cells is increasing. Cell type annotation of single-cell datasets is often performed based on known marker genes, integration with reference atlases, or models trained on a subset of genes (69). For different species,

these analyses are limited to orthologous genes that either must be specified or inferred with sophisticated tools, such as SAMap or SATURN (31, 70, 71). For scATAC-seq data, this usually requires relying on multiome data with both gene expression and ATAC-seq modalities. Modeling regulatory sequences circumvents the need for an additional gene expression modality or known orthologs, once a model has been trained on cell types from an annotated reference species atlas. Hence, enhancer codes identified by DL models provide a means to compare cell types solely on the basis of genome sequence and chromatin accessibility. Across all human, mouse, and chicken models, the prediction scores, nucleotide-level attribution scores, and motif combinations for MGLs, ASTs, OLs, and various interneuron types led to a high-confidence “consensus” level explanation of their enhancer code. Training a CREsted model on a combined set of mouse and human regions can further improve how well the model generalizes to unseen regions (fig. S57), in line with observations by other studies (14).

The excitatory neurons, as a broad class, share characteristic enhancer codes between species. However, at the subclass level, most of the excitatory neurons show one:many or many:many similarities between mammals and birds based on their transcriptome and enhancer codes. Putative homologies between regions of the vertebrate pallium have traditionally been assessed on the basis of their developmental origin or similarity in brain circuitry (19, 21, 72–74). On the basis of enhancer codes, we observe similarities between avian and mammalian excitatory neurons that do not fully agree with previously proposed evolutionary models based on shared developmental territories, nor with models based on vertebrate brain circuitry [see supplementary text (39)] (20, 75). For example, avian mesopallial GLU-7 and mammalian deep-layer L6 CT neurons share enhancer codes but are derived from different pallial fields during development (21). Also, the correspondence between GLU-1 neurons, which are distributed across the chicken nido- and hyperpallium and the mammalian neocortical upper layer, piriform cortex, and amygdala, does not strictly agree with these developmental models. Although many of the cell type correspondences that we identified share the same projection class, tracing of brain circuits in pigeons suggests that mesopallial neurons only project inside the telencephalon, which does not agree with the observed similarity based on enhancer codes between mesopallial GLU-7 and mammalian L6 CT neurons (19, 72). Despite these disagreements, the similarities in enhancer codes provide further support for two ancestral lineages of excitatory neurons that originated in the amniote ancestor or earlier (22, 76, 77): one

lineage comprising L6 CT/L6b, SUB CT, CLA, and GLU-7 neurons and the other comprising L2/3 IT, ENT, PIR, amygdala, and GLU-1 neurons. Differences in projection classes and developmental origins between mammalian and avian pallial neurons can likely be attributed to duplication, diversification, and co-option of enhancer codes and GRNs, as has previously been suggested for the olfactory pathway neurons (77).

Co-option and deletion of genomic enhancers constitute main drivers of cell type divergence (78). Supporting previous findings, we observed that homologous cell types, such as MGLs, have diverged transcriptomes between species but maintain conserved enhancer codes (fig. S58) (24, 79). Hence, our models provide powerful tools for studying how enhancer turnover relates to transcriptomic divergence. As an example, changes affecting OL enhancer code motifs in a predicted enhancer region near *Kirrel3* coincide with its expression gain and loss in mouse and chicken OLs, respectively (fig. S59). Although our approaches focus on conserved TF combinations and enhancer codes, they can be extended to study divergence of TF expression and TF combinations. This will provide insights about the extent to which adaptation, neutral evolution, and drift lead to changes of TF expression and enhancer codes between homologous cell types. For example, the expression of neocortical marker TFs, such as SATB2, BCL11B, FEZF2, and TBRI, does not match predicted mammal and chicken cell type correspondences [fig. S60, Fig. 6B, and supplementary text (39)].

Our study shows that sequence-based DL models can be used to robustly infer the identity of cell types and their characteristic enhancer codes directly from accessible genomic sequences across amniote species. In contrast to gene expression, enhancer codes can be directly traced across the genomes of related species to inform about evolutionary conservation. This provides a means to study cell type evolution through changes in candidate enhancers and the impact of genomic variants. DL models, as presented in this work, can be used to narrow down nucleotide changes that are crucial for changes in cell type specificity, as has been previously shown (10, 15). In past studies, we have verified that enhancer codes for melanoma states are conserved between mammalian and zebrafish cell lines and can identify variants in the genomes of melanoma patients (15, 80). The models presented in this study provide useful tools to study the impact of genomic variants and their association with mental or cognitive traits and disorders (81) and to design synthetic enhancers across the spectrum of brain cell types (82). Ultimately, models that learn the genomic regulatory code hold the potential to screen genomes and annotate genomic loci by cell

type specificity or to investigate the presence or absence of specific cell types or cell states in any species.

Materials and methods summary

We generated multiome (scRNA-seq and scATAC-seq) and spatially resolved transcriptomics data from the chicken telencephalon. Mouse and human brain multiome data were collected from previous studies. Data from all three species were analyzed through transcriptomic profiling from scRNA-seq and through sequence-based DL models trained on scATAC-seq data that predict chromatin accessibility per cell type from sequence. We compared cell types based on their transcriptomes using SAMap and based on their enhancer codes using cross-species predictions, nucleotide contribution score correlation, and learned motif similarity. In vivo confirmation of chicken enhancer activity and specificity for matching mouse cell types was performed by injection of AAV carrying an enhancer driving a GFP reporter gene. Imaging of brain slices costained with cell type markers revealed the specificity of the tested enhancers. See the supplementary materials for complete materials and methods.

REFERENCES AND NOTES

- S. Heinz *et al.*, Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589 (2010). doi: [10.1016/j.molcel.2010.05.004](https://doi.org/10.1016/j.molcel.2010.05.004); pmid: [20513432](https://pubmed.ncbi.nlm.nih.gov/20513432/)
- E. O. Mazzoni *et al.*, Synergistic binding of transcription factors to cell-specific enhancers programs motor neuron identity. *Nat. Neurosci.* **16**, 1219–1227 (2013). doi: [10.1038/nn.3467](https://doi.org/10.1038/nn.3467); pmid: [23872598](https://pubmed.ncbi.nlm.nih.gov/23872598/)
- D. Arendt, P. Y. Bertucci, K. Achim, J. M. Musser, Evolution of neuronal types and families. *Curr. Opin. Neurobiol.* **56**, 144–152 (2019). doi: [10.1016/j.conb.2019.01.022](https://doi.org/10.1016/j.conb.2019.01.022); pmid: [30826503](https://pubmed.ncbi.nlm.nih.gov/30826503/)
- C. Bravo González-Blas *et al.*, SCENIC+: Single-cell multiomic inference of enhancers and gene regulatory networks. *Nat. Methods* **20**, 1355–1367 (2023). doi: [10.1038/s41592-023-01938-4](https://doi.org/10.1038/s41592-023-01938-4); pmid: [37443338](https://pubmed.ncbi.nlm.nih.gov/37443338/)
- K. Kamimoto *et al.*, Dissecting cell identity via network inference and in silico gene perturbation. *Nature* **614**, 742–751 (2023). doi: [10.1038/s41586-022-05688-9](https://doi.org/10.1038/s41586-022-05688-9); pmid: [36755098](https://pubmed.ncbi.nlm.nih.gov/36755098/)
- G. Novakovsky, N. Dexter, M. W. Libbrecht, W. W. Wasserman, S. Mostafavi, Obtaining genetics insights from deep learning via explainable artificial intelligence. *Nat. Rev. Genet.* **24**, 125–137 (2023). doi: [10.1038/s41576-022-00532-2](https://doi.org/10.1038/s41576-022-00532-2); pmid: [36192604](https://pubmed.ncbi.nlm.nih.gov/36192604/)
- G. Eraslan, Ž. Avsec, J. Gagneur, F. J. Theis, Deep learning: New computational modelling techniques for genomics. *Nat. Rev. Genet.* **20**, 389–403 (2019). doi: [10.1038/s41576-019-0122-6](https://doi.org/10.1038/s41576-019-0122-6); pmid: [30971806](https://pubmed.ncbi.nlm.nih.gov/30971806/)
- I. M. Kaplow *et al.*, 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); pmid: [37104615](https://pubmed.ncbi.nlm.nih.gov/37104615/)
- R. K. Kawaguchi *et al.*, Learning single-cell chromatin accessibility profiles using meta-analytic marker genes. *Brief. Bioinform.* **24**, bbac541 (2023). doi: [10.1093/bib/bbac541](https://doi.org/10.1093/bib/bbac541); pmid: [36549922](https://pubmed.ncbi.nlm.nih.gov/36549922/)
- J. Janssens *et al.*, Decoding gene regulation in the fly brain. *Nature* **601**, 630–636 (2022). doi: [10.1038/s41586-021-04262-z](https://doi.org/10.1038/s41586-021-04262-z); pmid: [34987221](https://pubmed.ncbi.nlm.nih.gov/34987221/)
- C. Bravo González-Blas *et al.*, Single-cell spatial multi-omics and deep learning dissect enhancer-driven gene regulatory networks in liver zonation. *Nat. Cell Biol.* **26**, 153–167 (2024). doi: [10.1038/s41556-023-01316-4](https://doi.org/10.1038/s41556-023-01316-4); pmid: [38182825](https://pubmed.ncbi.nlm.nih.gov/38182825/)
- Ž. Avsec *et al.*, Base-resolution models of transcription-factor binding reveal soft motif syntax. *Nat. Genet.* **53**, 354–366 (2021). doi: [10.1038/s41588-021-00782-6](https://doi.org/10.1038/s41588-021-00782-6); pmid: [33603233](https://pubmed.ncbi.nlm.nih.gov/33603233/)
- S. Zu *et al.*, Single-cell analysis of chromatin accessibility in the adult mouse brain. *Nature* **624**, 378–389 (2023). doi: [10.1038/s41586-023-06824-9](https://doi.org/10.1038/s41586-023-06824-9); pmid: [38092917](https://pubmed.ncbi.nlm.nih.gov/38092917/)
- N. R. Zemke *et al.*, Conserved and divergent gene regulatory programs of the mammalian neocortex. *Nature* **624**, 390–402 (2023). doi: [10.1038/s41586-023-06819-6](https://doi.org/10.1038/s41586-023-06819-6); pmid: [38092918](https://pubmed.ncbi.nlm.nih.gov/38092918/)
- L. Minnoye *et al.*, Cross-species analysis of enhancer logic using deep learning. *Genome Res.* **30**, 1815–1834 (2020). doi: [10.1101/gr.260844.120](https://doi.org/10.1101/gr.260844.120); pmid: [32732264](https://pubmed.ncbi.nlm.nih.gov/32732264/)
- M. E. Wirthlin *et al.*, Vocal learning-associated convergent evolution in mammalian proteins and regulatory elements. *Science* **383**, eabn3263 (2024). doi: [10.1126/science.abn3263](https://doi.org/10.1126/science.abn3263); pmid: [38422184](https://pubmed.ncbi.nlm.nih.gov/38422184/)
- M. A. Tosches, From Cell Types to an Integrated Understanding of Brain Evolution: The Case of the Cerebral Cortex. *Annu. Rev. Cell Dev. Biol.* **37**, 495–517 (2021). doi: [10.1146/annurev-cellbio-120319-112654](https://doi.org/10.1146/annurev-cellbio-120319-112654); pmid: [34416113](https://pubmed.ncbi.nlm.nih.gov/34416113/)
- L. Pessoa, L. Medina, P. R. Hof, E. Desfilis, Neural architecture of the vertebrate brain: Implications for the interaction between emotion and cognition. *Neurosci. Biobehav. Rev.* **107**, 296–312 (2019). doi: [10.1016/j.neubiorev.2019.09.021](https://doi.org/10.1016/j.neubiorev.2019.09.021); pmid: [31541638](https://pubmed.ncbi.nlm.nih.gov/31541638/)
- M. Stacho *et al.*, A cortex-like canonical circuit in the avian forebrain. *Science* **369**, eabc5534 (2020). doi: [10.1126/science.abc5534](https://doi.org/10.1126/science.abc5534); pmid: [32973004](https://pubmed.ncbi.nlm.nih.gov/32973004/)
- S. D. Briscoe, C. W. Ragsdale, Homology, neocortex, and the evolution of developmental mechanisms. *Science* **362**, 190–193 (2018). doi: [10.1126/science.aau3711](https://doi.org/10.1126/science.aau3711); pmid: [30309947](https://pubmed.ncbi.nlm.nih.gov/30309947/)
- L. Puelles *et al.*, in *Evolution of Nervous Systems*, J. H. Kaas, Ed. (Elsevier, ed. 2, 2017), pp. 519–555. doi: [10.1016/B978-0-12-804042-3.00014-2](https://doi.org/10.1016/B978-0-12-804042-3.00014-2)
- M. A. Tosches *et al.*, Evolution of pallium, hippocampus, and cortical cell types revealed by single-cell transcriptomics in reptiles. *Science* **360**, 881–888 (2018). doi: [10.1126/science.aar4237](https://doi.org/10.1126/science.aar4237); pmid: [29724907](https://pubmed.ncbi.nlm.nih.gov/29724907/)
- B. M. Colquitt, D. P. Merullo, G. Konopka, T. F. Roberts, M. S. Brainard, Cellular transcriptomics reveals evolutionary identities of songbird vocal circuits. *Science* **371**, eabd9704 (2021). doi: [10.1126/science.abd9704](https://doi.org/10.1126/science.abd9704); pmid: [33574185](https://pubmed.ncbi.nlm.nih.gov/33574185/)
- T. E. Bakken *et al.*, 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); pmid: [34616062](https://pubmed.ncbi.nlm.nih.gov/34616062/)
- Y. E. Li *et al.*, 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); pmid: [34616068](https://pubmed.ncbi.nlm.nih.gov/34616068/)
- Z. Yao *et al.*, A high-resolution transcriptomic and spatial atlas of cell types in the whole mouse brain. *Nature* **624**, 317–332 (2023). doi: [10.1038/s41586-023-06812-z](https://doi.org/10.1038/s41586-023-06812-z); pmid: [38092916](https://pubmed.ncbi.nlm.nih.gov/38092916/)
- K. Siletti *et al.*, Transcriptomic diversity of cell types across the adult human brain. *Science* **382**, eadd7046 (2023). doi: [10.1126/science.add7046](https://doi.org/10.1126/science.add7046)
- Y. E. Li *et al.*, A comparative atlas of single-cell chromatin accessibility in the human brain. *Science* **382**, eadf7044 (2023). doi: [10.1126/science.adf7044](https://doi.org/10.1126/science.adf7044); pmid: [37824643](https://pubmed.ncbi.nlm.nih.gov/37824643/)
- V. Kleshchevnikov *et al.*, Cell2location maps fine-grained cell types in spatial transcriptomics. *Nat. Biotechnol.* **40**, 661–671 (2022). doi: [10.1038/s41587-021-01139-4](https://doi.org/10.1038/s41587-021-01139-4); pmid: [35027729](https://pubmed.ncbi.nlm.nih.gov/35027729/)
- S. Ma *et al.*, Molecular and cellular evolution of the primate dorsolateral prefrontal cortex. *Science* **377**, eabg7257 (2022). doi: [10.1126/science.abb07257](https://doi.org/10.1126/science.abb07257); pmid: [36007006](https://pubmed.ncbi.nlm.nih.gov/36007006/)
- A. J. Tarashansky *et al.*, Mapping single-cell atlases throughout Metazoa unravels cell type evolution. *eLife* **10**, e66747 (2021). doi: [10.7554/eLife.66747](https://doi.org/10.7554/eLife.66747); pmid: [33944782](https://pubmed.ncbi.nlm.nih.gov/33944782/)
- A. Chen *et al.*, Spatiotemporal transcriptomic atlas of mouse organogenesis using DNA nanoball-patterned arrays. *Cell* **185**, 1777–1792.e21 (2022). doi: [10.1016/j.cell.2022.04.003](https://doi.org/10.1016/j.cell.2022.04.003); pmid: [35512705](https://pubmed.ncbi.nlm.nih.gov/35512705/)
- S. Poovathingal *et al.*, Nova-ST: Nano-patterned ultra-dense platform for spatial transcriptomics. *Cell Rep. Methods* **4**, 100831 (2024). doi: [10.1016/j.crmeth.2024.100831](https://doi.org/10.1016/j.crmeth.2024.100831)
- A. Reiner, A new avian brain nomenclature: Why, how and what. *Brain Res. Bull.* **66**, 317–331 (2005). doi: [10.1016/j.brainresbull.2005.05.007](https://doi.org/10.1016/j.brainresbull.2005.05.007); pmid: [16144608](https://pubmed.ncbi.nlm.nih.gov/16144608/)
- G. Gedman *et al.*, As above, so below: Whole transcriptome profiling demonstrates strong molecular similarities between avian dorsal and ventral pallial subdivisions. *J. Comp. Neurol.* **529**, 3222–3246 (2021). doi: [10.1002/cne.25159](https://doi.org/10.1002/cne.25159); pmid: [33871048](https://pubmed.ncbi.nlm.nih.gov/33871048/)
- B. Zaremba *et al.*, Developmental origins and evolution of pallial cell types and structures in birds. *Science* **387**, eadp5182 (2025). doi: [10.1126/science.adp5182](https://doi.org/10.1126/science.adp5182)
- E. Rueda-Alaría *et al.*, Evolutionary convergence of sensory circuits in the pallium of amniotes. *Science* **387**, eadp3411 (2025). doi: [10.1126/science.adp3411](https://doi.org/10.1126/science.adp3411)
- M. Bergslund, M. Werne, M. Malewitz, T. Perlmann, J. Muhr, The establishment of neuronal properties is controlled by Sox4 and Sox11. *Genes Dev.* **20**, 3475–3486 (2006). doi: [10.1101/gad.403406](https://doi.org/10.1101/gad.403406); pmid: [17182872](https://pubmed.ncbi.nlm.nih.gov/17182872/)
- See the supplementary materials.
- N. Almeida *et al.*, Employing core regulatory circuits to define cell identity. *EMBO J.* **40**, e106785 (2021). doi: [10.15252/embj.2020106785](https://doi.org/10.15252/embj.2020106785); pmid: [33934382](https://pubmed.ncbi.nlm.nih.gov/33934382/)
- E. H. Davidson, D. H. Erwin, Gene regulatory networks and the evolution of animal body plans. *Science* **311**, 796–800 (2006). doi: [10.1126/science.1113832](https://doi.org/10.1126/science.1113832)
- D. Arendt *et al.*, The origin and evolution of cell types. *Nat. Rev. Genet.* **17**, 744–757 (2016). doi: [10.1038/nrg.2016.127](https://doi.org/10.1038/nrg.2016.127); pmid: [27818507](https://pubmed.ncbi.nlm.nih.gov/27818507/)
- S. M. Lundberg, S.-I. Lee, in *Advances in Neural Information Processing Systems*, I. Guyon *et al.*, Eds. (Curran Associates, Inc., 2017), pp. 4765–4774.
- A. Shrikumar *et al.*, Proprietary Activation Differences. *arXiv:1704.02685* [cs.CV] (2019).
- D. R. Kelley, J. Snoek, J. L. Rinin, Bassett: Learning the regulatory code of the accessible genome with deep convolutional neural networks. *Genome Res.* **26**, 990–999 (2016). doi: [10.1101/gr.200535.115](https://doi.org/10.1101/gr.200535.115); pmid: [2797224](https://pubmed.ncbi.nlm.nih.gov/2797224)
- J. Zhou, O. G. Troyanskaya, Predicting effects of noncoding variants with deep learning-based sequence model. *Nat. Methods* **12**, 931–934 (2015). doi: [10.1038/nmeth.3547](https://doi.org/10.1038/nmeth.3547); pmid: [26301843](https://pubmed.ncbi.nlm.nih.gov/26301843/)
- K. A. Sauter *et al.*, The function of the conserved regulatory element within the second intron of the mammalian *Csf1r* locus. *PLOS ONE* **8**, e54935 (2013). doi: [10.1371/journal.pone.0054935](https://doi.org/10.1371/journal.pone.0054935); pmid: [23883005](https://pubmed.ncbi.nlm.nih.gov/23883005/)
- Ž. Avsec *et al.*, Effective gene expression prediction from sequence by integrating long-range interactions. *Nat. Methods* **18**, 1196–1203 (2021). doi: [10.1038/s41592-021-01252-x](https://doi.org/10.1038/s41592-021-01252-x); pmid: [34608324](https://pubmed.ncbi.nlm.nih.gov/34608324/)
- A. Pampliega *et al.*, ChromoBPNet: Bias factorized, base-resolution deep learning models of chromatin accessibility reveal cis-regulatory sequence syntax, transcription factor footprints and regulatory variants. *bioRxiv* 2024.12.25.630221 [Preprint] (2025); <https://doi.org/10.1101/2024.12.25.630221>
- J.-M. Baizabal *et al.*, The Epigenetic State of PRDM16-Regulated Enhancers in Radial Glia Controls Cortical Neuron Position. *Neuron* **98**, 945–962.e8 (2018). doi: [10.1016/j.neuron.2018.04.033](https://doi.org/10.1016/j.neuron.2018.04.033); pmid: [2977941](https://pubmed.ncbi.nlm.nih.gov/2977941)
- A. Shrikumar *et al.*, Technical Note on Transcription Factor Motif Discovery from Importance Scores (TF-MoDISCo) version 0.5.6.5. *arXiv:1811.00416* [cs.LG] (2020).
- S. Gupta, J. A. Stamatoyannopoulos, T. L. Bailey, W. S. Noble, Quantifying similarity between motifs. *Genome Biol.* **8**, R24 (2007). doi: [10.1186/gb-2007-8-2-24](https://doi.org/10.1186/gb-2007-8-2-24); pmid: [17324271](https://pubmed.ncbi.nlm.nih.gov/17324271/)
- T. L. Bailey, J. Johnson, C. E. Grant, W. S. Noble, The MEME Suite. *Nucleic Acids Res.* **43**, W39–W49 (2015). doi: [10.1093/nar/gkv416](https://doi.org/10.1093/nar/gkv416); pmid: [25953851](https://pubmed.ncbi.nlm.nih.gov/25953851/)
- N. Hecker, N. Kempynck, D. Mauduit, S. Aerts, Additional data to ‘Enhancer-driven cell type comparison reveals similarities between the mammalian and bird pallium’. *Zenodo* (2024); <https://doi.org/10.5281/zenodo.13349532>
- N. Kempynck *et al.*, CREsted: Cis Regulatory Element Sequence Training, Explanation, and Design, version 1.0.0. *Zenodo* (2024); <https://doi.org/10.5281/zenodo.13320756>
- K. C. Allaway *et al.*, Genetic and epigenetic coordination of cortical interneuron development. *Nature* **597**, 693–697 (2021). doi: [10.1038/s41586-021-03933-1](https://doi.org/10.1038/s41586-021-03933-1); pmid: [34552240](https://pubmed.ncbi.nlm.nih.gov/34552240/)
- K. Takahashi, F.-C. Liu, K. Hirokawa, H. Takahashi, Expression of *Foxp2*, a gene involved in speech and language, in the developing and adult striatum. *J. Neurosci. Res.* **73**, 61–72 (2003). doi: [10.1002/jnr.10638](https://doi.org/10.1002/jnr.10638); pmid: [12815709](https://pubmed.ncbi.nlm.nih.gov/12815709/)
- C. A. French *et al.*, An aetiological *Foxp2* mutation causes aberrant striatal activity and alters plasticity during skill learning. *Mol. Psychiatry* **17**, 1077–1085 (2012). doi: [10.1038/mp.2011.105](https://doi.org/10.1038/mp.2011.105); pmid: [21876543](https://pubmed.ncbi.nlm.nih.gov/21876543/)
- W. Enard *et al.*, A humanized version of *Foxp2* affects cortico-basal ganglia circuits in mice. *Cell* **137**, 961–971 (2009). doi: [10.1016/j.cell.2009.03.041](https://doi.org/10.1016/j.cell.2009.03.041); pmid: [19490899](https://pubmed.ncbi.nlm.nih.gov/19490899/)
- L. Medina, A. Abellán, E. Desfilis, Evolving Views on the Pallium. *Brain Behav. Evol.* **96**, 181–199 (2022). doi: [10.1159/000519260](https://doi.org/10.1159/000519260); pmid: [34657034](https://pubmed.ncbi.nlm.nih.gov/34657034/)

61. H. Peng *et al.*, Morphological diversity of single neurons in molecularly defined cell types. *Nature* **598**, 174–181 (2021). doi: [10.1038/s41586-021-03941-1](https://doi.org/10.1038/s41586-021-03941-1); pmid: [34616072](https://pubmed.ncbi.nlm.nih.gov/34616072/)
62. D. P. Leone, K. Srinivasan, B. Chen, E. Alcamo, S. K. McConnell, The determination of projection neuron identity in the developing cerebral cortex. *Curr. Opin. Neurobiol.* **18**, 28–35 (2008). doi: [10.1016/j.conb.2008.05.006](https://doi.org/10.1016/j.conb.2008.05.006); pmid: [18508260](https://pubmed.ncbi.nlm.nih.gov/18508260/)
63. K. Okamoto *et al.*, Specific AAV2/PHP.eB-mediated gene transduction of CA2 pyramidal cells via injection into the lateral ventricle. *Sci. Rep.* **13**, 323 (2023). doi: [10.1038/s41598-022-27372-8](https://doi.org/10.1038/s41598-022-27372-8); pmid: [36609635](https://pubmed.ncbi.nlm.nih.gov/36609635/)
64. K. Y. Chan *et al.*, Engineered AAVs for efficient noninvasive gene delivery to the central and peripheral nervous systems. *Nat. Neurosci.* **20**, 1172–1179 (2017). doi: [10.1038/nn.4593](https://doi.org/10.1038/nn.4593); pmid: [28671695](https://pubmed.ncbi.nlm.nih.gov/28671695/)
65. J. Hahn *et al.*, Evolution of neuronal cell classes and types in the vertebrate retina. *Nature* **624**, 415–424 (2023). doi: [10.1038/s41586-023-06638-9](https://doi.org/10.1038/s41586-023-06638-9); pmid: [38092908](https://pubmed.ncbi.nlm.nih.gov/38092908/)
66. P. Lyu *et al.*, Gene regulatory networks controlling temporal patterning, neurogenesis, and cell-fate specification in mammalian retina. *Cell Rep.* **37**, 109994 (2021). doi: [10.1016/j.celrep.2021.109994](https://doi.org/10.1016/j.celrep.2021.109994); pmid: [34788628](https://pubmed.ncbi.nlm.nih.gov/34788628/)
67. L. Qian *et al.*, Transcription factor neuromancer/TBX20 is required for cardiac function in *Drosophila* with implications for human heart disease. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 19833–19838 (2008). doi: [10.1073/pnas.0808705105](https://doi.org/10.1073/pnas.0808705105); pmid: [19074289](https://pubmed.ncbi.nlm.nih.gov/19074289/)
68. E. S. Wong *et al.*, 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); pmid: [33154111](https://pubmed.ncbi.nlm.nih.gov/33154111/)
69. L. Heumos *et al.*, Best practices for single-cell analysis across modalities. *Nat. Rev. Genet.* **24**, 550–572 (2023). doi: [10.1038/s41586-023-00586-w](https://doi.org/10.1038/s41586-023-00586-w); pmid: [37002403](https://pubmed.ncbi.nlm.nih.gov/37002403/)
70. Y. Song, Z. Miao, A. Brazma, I. Papathodorou, Benchmarking strategies for cross-species integration of single-cell RNA sequencing data. *Nat. Commun.* **14**, 6495 (2023). doi: [10.1038/s41467-023-41855-w](https://doi.org/10.1038/s41467-023-41855-w); pmid: [37838716](https://pubmed.ncbi.nlm.nih.gov/37838716/)
71. Y. Rosen *et al.*, Toward universal cell embeddings: Integrating single-cell RNA-seq datasets across species with SATURN. *Nat. Methods* **21**, 1492–1500 (2024). doi: [10.1038/s41592-024-02191-z](https://doi.org/10.1038/s41592-024-02191-z); pmid: [38366243](https://pubmed.ncbi.nlm.nih.gov/38366243/)
72. O. Güntürkün, R. Pusch, J. Rose, Why birds are smart. *Trends Cogn. Sci.* **28**, 197–209 (2024). doi: [10.1016/j.tics.2023.11.002](https://doi.org/10.1016/j.tics.2023.11.002); pmid: [38097447](https://pubmed.ncbi.nlm.nih.gov/38097447/)
73. H. J. Karten, The Organization of the Avian Telencephalon and Some Speculations on the Phylogeny of the Amniote Telencephalon. *Ann. N. Y. Acad. Sci.* **167**, 164–179 (1969). doi: [10.1111/j.1749-6632.1969.tb20442.x](https://doi.org/10.1111/j.1749-6632.1969.tb20442.x)
74. L. Puelles *et al.*, Pallial and subpallial derivatives in the embryonic chick and mouse telencephalon, traced by the expression of the genes Dlx-2, Emx-1, Nkx-2.1, Pax-6, and Tbr-1. *J. Comp. Neurol.* **424**, 409–438 (2000). doi: [10.1002/1096-9861\(20000828\)424:3<409::AID-CNE3>3.0.CO;2-7](https://doi.org/10.1002/1096-9861(20000828)424:3<409::AID-CNE3>3.0.CO;2-7); pmid: [10906711](https://pubmed.ncbi.nlm.nih.gov/10906711/)
75. S. D. Briscoe, C. B. Albertin, J. J. Rowell, C. W. Ragsdale, Neocortical Association Cell Types in the Forebrain of Birds and Alligators. *Curr. Biol.* **28**, 686–696.e6 (2018). doi: [10.1016/j.cub.2018.01.036](https://doi.org/10.1016/j.cub.2018.01.036); pmid: [29456143](https://pubmed.ncbi.nlm.nih.gov/29456143/)
76. J. Woych *et al.*, Cell-type profiling in salamanders identifies innovations in vertebrate forebrain evolution. *Science* **377**, eabb9186 (2022). doi: [10.1126/science.abp9186](https://doi.org/10.1126/science.abp9186); pmid: [36048957](https://pubmed.ncbi.nlm.nih.gov/36048957/)
77. F. Luzzati, A hypothesis for the evolution of the upper layers of the neocortex through co-option of the olfactory cortex developmental program. *Front. Neurosci.* **9**, 162 (2015). doi: [10.3389/fnins.2015.00162](https://doi.org/10.3389/fnins.2015.00162); pmid: [26029038](https://pubmed.ncbi.nlm.nih.gov/26029038/)
78. H. K. Long, S. L. Prescott, J. Wysocka, Ever-Changing Landscapes: Transcriptional Enhancers in Development and Evolution. *Cell* **167**, 1170–1187 (2016). doi: [10.1016/j.cell.2016.09.018](https://doi.org/10.1016/j.cell.2016.09.018); pmid: [27863239](https://pubmed.ncbi.nlm.nih.gov/27863239/)
79. L. Geirsdottir *et al.*, Cross-Species Single-Cell Analysis Reveals Divergence of the Primate Microglia Program. *Cell* **179**, 1609–1622.e16 (2019). doi: [10.1016/j.cell.2019.11.010](https://doi.org/10.1016/j.cell.2019.11.010); pmid: [31835035](https://pubmed.ncbi.nlm.nih.gov/31835035/)
80. Z. K. Atak *et al.*, Interpretation of allele-specific chromatin accessibility using cell state-aware deep learning. *Genome Res.* **31**, 1082–1096 (2021). doi: [10.1101/gr.260851.120](https://doi.org/10.1101/gr.260851.120); pmid: [33832990](https://pubmed.ncbi.nlm.nih.gov/33832990/)
81. K. M. Chen, A. K. Wong, O. G. Troyanskaya, J. Zhou, A sequence-based global map of regulatory activity for deciphering human genetics. *Nat. Genet.* **54**, 940–949 (2022). doi: [10.1038/s41588-022-01102-2](https://doi.org/10.1038/s41588-022-01102-2); pmid: [35817977](https://pubmed.ncbi.nlm.nih.gov/35817977/)
82. I. I. Taskiran *et al.*, Cell-type-directed design of synthetic enhancers. *Nature* **626**, 212–220 (2024). doi: [10.1038/s41586-023-06936-2](https://doi.org/10.1038/s41586-023-06936-2); pmid: [38086419](https://pubmed.ncbi.nlm.nih.gov/38086419/)
83. N. Kempynck, DeepBrain, Zenodo (2024); <https://doi.org/10.5281/zenodo.13349958>.

ACKNOWLEDGMENTS

We thank E. Seuntjens and V. Darras for helping with the chicken brain dissections; P. Vanderhaeghen for advice; and B. Zaremba, F. Garcia-Moreno, and H. Kaessmann for helpful comments on the manuscript. We further thank the VIB Bioimaging Core for help with the Stereo-seq data, K. Theunis for help with the Nova-ST data, and the Vlaams Supercomputer Center and T. Van de Woestyne from the VIB Data Core for computing infrastructure. We also thank C. Calatayud Aristoy, D. Daaboul, B. Lorente Echeverría, G. Marcassa, and D. Dascenco for their help and advice on virus productions and injections, imaging, and image analysis. Finally, we thank A. Pampari and A. Kundaje for early access and insights to the ChromBPNet framework. **Funding:** This study was supported by European Research Council advanced grant 101054387 (S.A.), Chan Zuckerberg Initiative grant DI2-0000000068 (S.A.), Research Foundation – Flanders (FWO) strategic basic research grant S005024N (S.A.), FWO senior postdoctoral fellowship 1273822N (N.H.), FWO PhD fellowship strategic basic research 1SH6J24N (N.K.), FWO PhD fellowship 11F1519N (C.B.G.-B.), European Molecular Biology Organization (EMBO) postdoctoral fellowship to ALTF 638-2023 (L.B.), EMBO scientific exchange grant 9231 (I.S.), FWO junior research project fundamental research G057121N (L.L.), and FWO Odysseus Award GOE9121N (L.L.). **Author contributions:** N.H. and N.K. are listed as co-first authors because they contributed equally to the manuscript. Computational analysis: N.H., N.K., D.A., I.S., C.B.G.-B., J.D.M.; Conceptualization: N.H., S.A.; Data collection, processing, and curation: N.H., G.H., N.K., C.B.G.-B., L.B., K.D.; Experiments and sample preparation: D.M., R.V., S.D., E.L., J.V., R.M., V.C., S.P.; Funding acquisition: S.A., N.H., N.K.; Methodology: N.H., N.K., S.A.; Resources: S.P., L.L., J.D.W., S.A.; Software implementation and testing: N.K., N.H., D.A., I.S.; Visualization: N.H., N.K., D.M., D.A., L.B.; Writing – original draft: N.H., N.K., S.A.; Writing – review & editing: N.H., N.K., D.M., C.B.G.-B., I.S., S.A. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** The weights of the DeepMouseBrain1 and 2, DeepHumanCortex1 and 2, and DeepChickenBrain1 models and the code for making predictions, getting contribution scores, and calculating cell type contribution score correlation are available at Github (<https://github.com/aertslab/DeepBrain>) and Zenodo ([83](https://zenodo.org/record/83)). A UCSC track hub of the scATAC profiles of all datasets is available at https://ucsctracks.aertslab.org/papers/brain_evo/hub.txt. Processed single-cell and spatially resolved transcriptomics datasets are available as Scope session at https://scope.aertslab.org/#/PalliumEvolution/*welcome. Additional data are available on Zenodo ([54](https://zenodo.org/record/54)), containing all the TF-MoDiSc0 pattern results, the code for the MPRA analysis, and a storage location for all model weights. Raw and processed data of the chicken telencephalon 10x Genomics Single Cell Multiome ATAC + Gene expression, chicken telencephalon spatially resolved transcriptomics (Stereo-seq and Nova-ST), and MPRA in BV2 cell lines that were generated for this study are available at the NCBI Gene Expression Omnibus (GEO): <https://www.ncbi.nlm.nih.gov/geo/>; GSE26230 (MPRA), GSE26231 (single-cell multiome), GSE26232 (Stereo-seq), and GSE275097 (Nova-ST). **License information:** Copyright © 2025 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.adp3957
Materials and Methods
Supplementary Text
Figs. S1 to S60
Tables S1 to S13
References (84–105)
MDAR Reproducibility Checklist
Submitted 27 March 2024; accepted 26 November 2024
10.1126/science.adp3957