Title: Relating enhancer genetic variation across mammals to complex phenotypes using machine learning

One Sentence Summary: A new machine learning-based approach associates enhancers with the evolution of brain size and behavior across mammals.

Authors: Irene M. Kaplow[1,3,*,#], Alyssa J. Lawler[2,3,4,*], Daniel E. Schäffer[1,*], Chaitanya Srinivasan[1], Morgan E. Wirthlin[1,3], BaDoi N. Phan[1,3,5], Xiaomeng Zhang[1,4], Kathleen Foley[6,7], Kavya Prasad[1], Ashley R. Brown[1], Zoonomia Consortium, Wynn K. Meyer[6], and Andreas R. Pfenning[1,2,3,#]

*These authors contributed equally to this work.

Affiliations: Carnegie Mellon University Departments of [1]Computational Biology and [2]Biology and [3]Neuroscience Institute; [4]Current affiliation: Broad Institute, Massachusetts Institute of Technology; [5]Medical Scientist Training Program, University of Pittsburgh School of Medicine; [6]Department of Biological Sciences, Lehigh University; [7]Current affiliation: College of Law, University of Iowa

[#]Address correspondences to apfenning@cmu.edu, ikaplow@cs.cmu.edu

Abstract:

Protein-coding differences between mammals often fail to explain phenotypic diversity, suggesting involvement of enhancers, often rapidly evolving regions that regulate gene expression. Identifying associations between enhancers and phenotypes is challenging because enhancer activity is context-dependent and may be conserved without much sequence conservation. We developed TACIT (Tissue-Aware Conservation Inference Toolkit) to associate open chromatin regions (OCRs) with phenotypes using predictions in hundreds of mammalian genomes from machine learning models trained to learn tissue-specific regulatory codes. Applying TACIT for motor cortex and parvalbumin-positive interneurons to neurological phenotypes revealed dozens of new OCR-phenotype associations. Many

associated OCRs were near relevant genes, including brain size-associated OCRs near genes mutated in microcephaly or macrocephaly. Our work creates a forward genomics foundation for identifying candidate enhancers associated with phenotype evolution.

Main Text:

INTRODUCTION

Much of the phenotypic diversity that exists across vertebrates is thought to have arisen from differences in how genes are expressed (*1*). Variation in phenotypes like vocal learning (*2*) and longevity (*3*) has been linked to patterns of gene expression within some of the most relevant brain regions and tissues, respectively. Thus, many genetic differences associated with the evolution of these, and other, complex phenotypes are likely in enhancers, distal *cis*-regulatory genomic elements that are bound by transcription factor (TF) proteins that regulate the expression of associated genes, often through cell type-specific activation (*4*, *5*). For example, limblessness in snakes is associated with sequence divergence and activity loss in a critical enhancer near the *Sonic hedgehog* gene (*6*), and mutations in orthologs of this enhancer are associated with polydactyly in humans, mice, and cats (*7*, *8*). Enhancer evolution has been found to be associated with a number of other complex phenotypes, including eyesight loss (*9*) as well as whisker, penile spine, and brain growth (*10*).

Recent advances facilitate identifying relationships between enhancer activity and phenotype evolution. Community genome sequencing efforts such as the Zoonomia Project have constructed assemblies for hundreds of species from diverse mammalian clades (*11*). Cactus multi-species whole-genome alignments and tools for extracting orthologs have vastly improved ortholog mapping for non-coding genomic regions (*12–14*). In addition, new phylogeny-aware statistical methods have been developed for identifying factors associated with the evolution of phenotypes (*15*, *16*).

Despite these successes, identifying enhancer-phenotype relationships is still a major challenge. Widely used methods to identify conservation and convergent evolution across orthologous genome sequences measure the extent to which the nucleotides within a given region align across species (*17–19*). While these approaches have led to some exciting findings (*9*, *20*), many enhancer sequences and transcription factor binding sites are under less sequence constraint than promoter and gene sequences (*21*, *22*). In fact, recent studies have shown that sequence conservation is not required for activity conservation at enhancer orthologs (*23*, *24*) and can occur when enhancer activity is not conserved in a tissue of interest (*25*), so nucleotide sequence conservation at enhancers is sometimes an insufficient proxy for enhancer activity conservation.

Here we present a new method for identifying enhancer-phenotype associations, in which we trace enhancer activity evolution using predicted open chromatin in a tissue or cell type of interest as a proxy for enhancer function. Previously, we and others have demonstrated that the sequence patterns associated with enhancer activity in multiple tissues are highly conserved across mammals by showing that machine learning models that use DNA sequence to predict enhancer activity in a tissue of interest in one species can accurately predict clade-specific and tissue-specific enhancer activity in species from different mammalian clades (*25*,*27–29*). We integrate machine learning-based predictions of enhancer function with other comparative genomics advances (*11*, *15*, *16*) in a new framework called the Tissue-Aware Conservation Inference Toolkit (TACIT) for identifying candidate enhancers associated with the evolution of phenotypes. We use sequences underlying open chromatin regions (OCRs) from a small number of species in a tissue or cell type of interest to train convolutional neural networks (CNNs) that predict the probability of OCR ortholog open chromatin in those tissues/cell types at the orthologous sequences in up to 222 mammalian genomes (*11*). We then use these predictions to link OCRs to specific mammalian phenotypes while accounting for phylogeny (**Fig. 1**). We applied our approach to multiple phenotypes, including brain size, solitary and group living, and vocal learning, and identified both motor cortex tissue and motor cortex parvalbumin-positive (PV+) interneuron OCRs associated with these phenotypes that are near relevant genes. Our approach can be applied to any phenotype with open

chromatin data available from a relevant tissue or cell type in at least two species. It is therefore broadly applicable to a variety of tissue, phenotype, and species combinations.

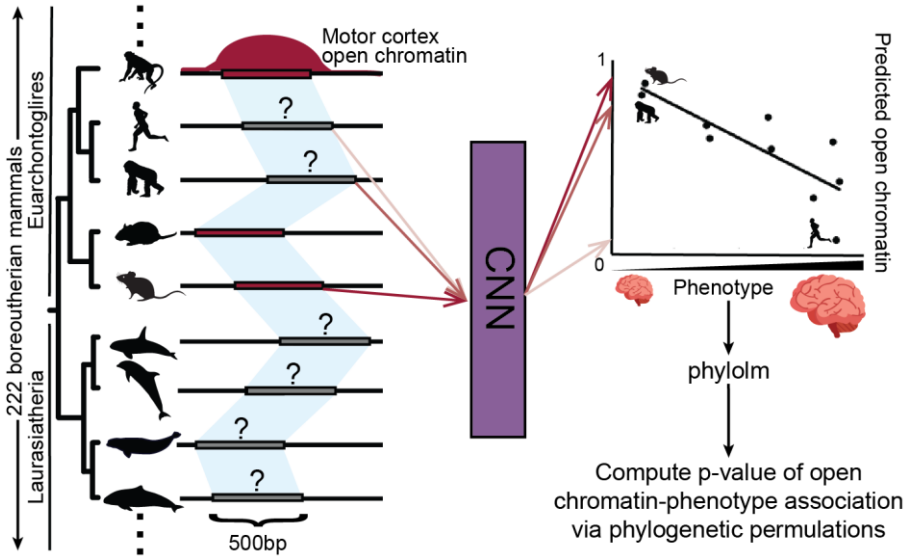

Figure 1: Overview of TACIT.
We train a CNN using sequences underlying OCRs and non-OCRs to predict open chromatin in a tissue or cell type of interest and then use the CNN to predict open chromatin in that tissue or cell type in hundreds of genomes from Zoonomia. We associate our predictions with phenotypes using phylolm and then quantify the significance of the association using an empirical p-value from phylogenetic permutations. Animal silhouettes were made by Michael Keesey, Daniel Jaron, Ryan Cupo, Steven Traver, and Chris Huh (license: https://creativecommons.org/licenses/by-sa/3.0/); were downloaded from PhyloPic; and were not modified.

RESULTS

Convolutional neural networks accurately predict open chromatin status of OCR orthologs

We applied TACIT to two tissues with open chromatin data from more than two species – motor cortex and liver – as well as a tissue and a cell type with data from only two species – retina and motor cortex PV+ interneurons. We used OCRs instead of other enhancer activity measures, such as H3K27ac ChIP-seq regions, because OCRs tend to have a concentration of TF motifs near their summits and be hundreds instead of thousands of base pairs long, allowing our model to focus on sequences likely to be involved in enhancer activity and allowing us to easily map regions in species whose assemblies have

short scaffolds (*14*). We chose tissues and cell types that would demonstrate specificity in dissimilar tissues (brain versus liver) and have relationships with complex phenotypes of interest, including brain size, social behavior, and vocal learning. For tissues with more than two species, we trained CNNs to predict whether a region is an OCR or a non-OCR ortholog of an OCR, as described in our previous work (*25*).

Since we are the first to train machine learning models for open chromatin prediction in motor cortex (we and others have shown that the liver regulatory code is conserved across species (*25*, *27*)), we first trained CNNs using only house mouse sequences and found that the CNNs successfully predicted clade-specific OCRs and non-OCRs (high "lineage-specific OCR accuracy," AUC > 0.70 and AUPRC/NPV-Spec. > 0.65 for all metrics) as well as tissue-specific OCRs and non-OCRs (high "tissue-specific OCR accuracy," AUC > 0.65 and AUPRC/NPV-Spec. > fraction of examples in smaller class for all metrics); in addition, when comparing average OCR ortholog predictions across species, predictions had the expected negative correlation with distance from the species in which the OCRs were assayed (high "phylogeny-matching correlations," mean Pearson correlation < -0.70 and mean Spearman correlation < -0.45) (**Figs. S1A,D,G,J,M,P, Table S1**) (*25*). We next trained multi-species CNNs for motor cortex and liver using all of our data – *Mus musculus* (Glires clade), *Macaca mulatta* (Euarchonta clade), and *Rattus norvegicus* (Glires clade) for both tissues as well as *Rousettus aegyptiacus* (Laurasiatheria clade) for motor cortex and *Bos taurus* (Laurasiatheria clade) and *Sus scrofa* (Laurasiatheria clade) for liver – and found that the models achieved high lineage- and tissue-specific OCR accuracy (AUC > 0.8, AUPRC/NPV-Spec. > fraction of examples in smaller class for all metrics) as well as phylogeny-matching correlations (mean Pearson correlation < -0.95 and mean Spearman correlation < -0.75) (**Fig. 2**, **Figs. S2A,D,G, Fig. S3, Tables S2-3**). We then used the multi-species motor cortex CNN to make predictions at motor cortex OCR orthologs in 222 diverse boreoeutherian mammal genomes from Zoonomia, where we limited ourselves to boreoeutherians because we did not have open chromatin data from species in other clades. To further evaluate the reliability of our predictions, we clustered the species hierarchically with predictions as features and found that the cluster hierarchy was

similar to the phylogenetic tree, with all but a few species clustering correctly by clade (**Fig. S4, Supplementary Text**) (*26*).

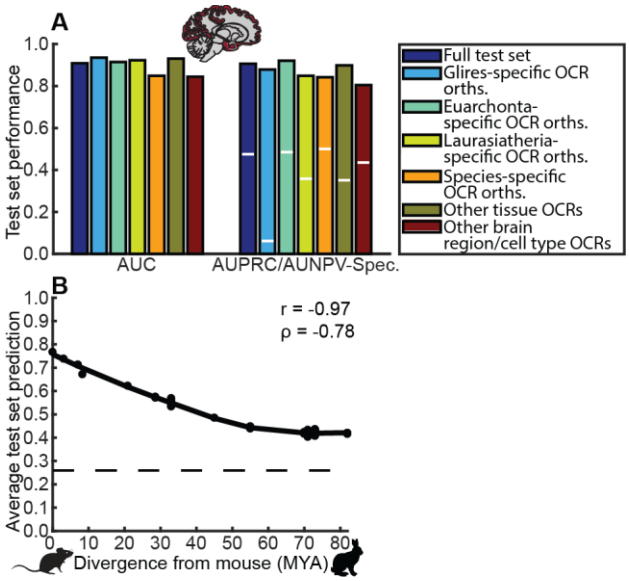

Figure 2: Motor cortex multi-species CNN performance.
A shows the area under the ROC curve (AUROC) and the area under the precision-recall (if more negatives than positives)/negative predictive value-specificity (if more positives than negatives) curve (AUPRC/NPV-Spec.) for the full test set, clade-specific OCRs and non-OCRs, and shared versus tissue/brain region-specific OCRs and non-OCRs for the multi-species motor cortex CNNs. **B** shows the negative relationship between the average house mouse OCR ortholog multi-species motor cortex open chromatin predicted probabilities for Glires species and the millions of years ago (MYA) when each species diverged from house mouse.

Since no previous study has trained PV+ interneuron or retinal enhancer activity prediction models for predicting enhancer activity in species not used for training (*25,27–29*), we needed to investigate whether the PV+ interneuron and retinal regulatory codes are sufficiently conserved for accurately predicting open chromatin of OCR orthologs. We did this by running motif discovery on open chromatin datasets from each species for which data was available. For each of PV+ interneurons and retina, we found motifs for many of the same TFs in both species, and some of these TFs are known to be involved in PV+ interneurons and retina, respectively (**Supplementary Text, Supplementary Website**) (*26*).

Because we had PV+ interneuron and retina data from only two species – *Mus musculus* and *Homo sapiens* (Euarchonta clade) – we did not have sufficient non-OCR orthologs of OCRs to train CNNs, so we developed a new approach to constructing negative sets for these cases: We combined a

large number of random regions of the genome with the same G/C-content as the positives with OCRs from other cell types or tissues, two negative sets that provided adequate performance for all of our metrics in our previous work (**Methods**) (*25*). To ensure that CNNs could make accurate predictions in species not used for training in our tissues and cell types, we first trained CNNs using only house mouse sequences and found that they achieved high lineage-specific OCR accuracy (AUC > 0.85 and AUPRC/NPV-Spec. > 0.60) as well as phylogeny-matching correlations (mean Pearson correlation < -0.65, mean Spearman correlation < -0.40 for retina and PV+ interneurons) for house mouse sequences (**Figs. S1B,C,E,F,H,I,K,L,N,O,Q,R, Tables S4-5**). The PV+ interneuron CNNs also achieved strong performance on human sequences (AUC > 0.70 and AUPRC/NPV-Spec. > fraction of examples in minority class for all criteria), where no human sequences were used in training as well as high tissue-specific OCR accuracy (AUC > 0.75 and AUPRC/NPV-Spec. > fraction of examples in minority class for all criteria), while the house mouse-trained retina CNNs did not work as well on human-specific OCRs and non-OCRs and liver non-retina OCRs. We then trained CNNs using sequences from both house mouse and human, and both the PV+ and retina CNNs achieved strong performance for all criteria (AUC > 0.70 and AUPRC/NPV-Spec. > fraction of examples in minority class for all criteria, mean Pearson correlation < -0.60, mean Spearman correlation < -0.40) (**Figs. 3A-D**, **Figs. S2B,C,E,F,H,I, Tables S6-7**).

To evaluate if our bulk tissue models were learning sequences relevant to the tissues in which they were trained, we interpreted what they had learned (**Methods**). Specifically, we computed the CNNs' per-nucleotide importance scores, which indicate the extent to which the CNN prioritizes the presence or absence of each nucleotide at each position (*30*, *31*). We found that our CNNs seemed to have learned sequence patterns that are similar to motifs of TFs that are known to be involved in motor cortex and liver, such as MEF2C for motor cortex (*32*, *33*) and HNF4A (*34*, *35*) for liver, as well as sequence patterns that do not match any known TF motif (**Supplementary Text**, **Figs. S5-7**) (*26*). We then examined a specific retina OCR near the retina TF *Otx2*, where the OCR's orthologs in subterranean mammals were previously shown to have a faster relative evolutionary rate than its orthologs in other

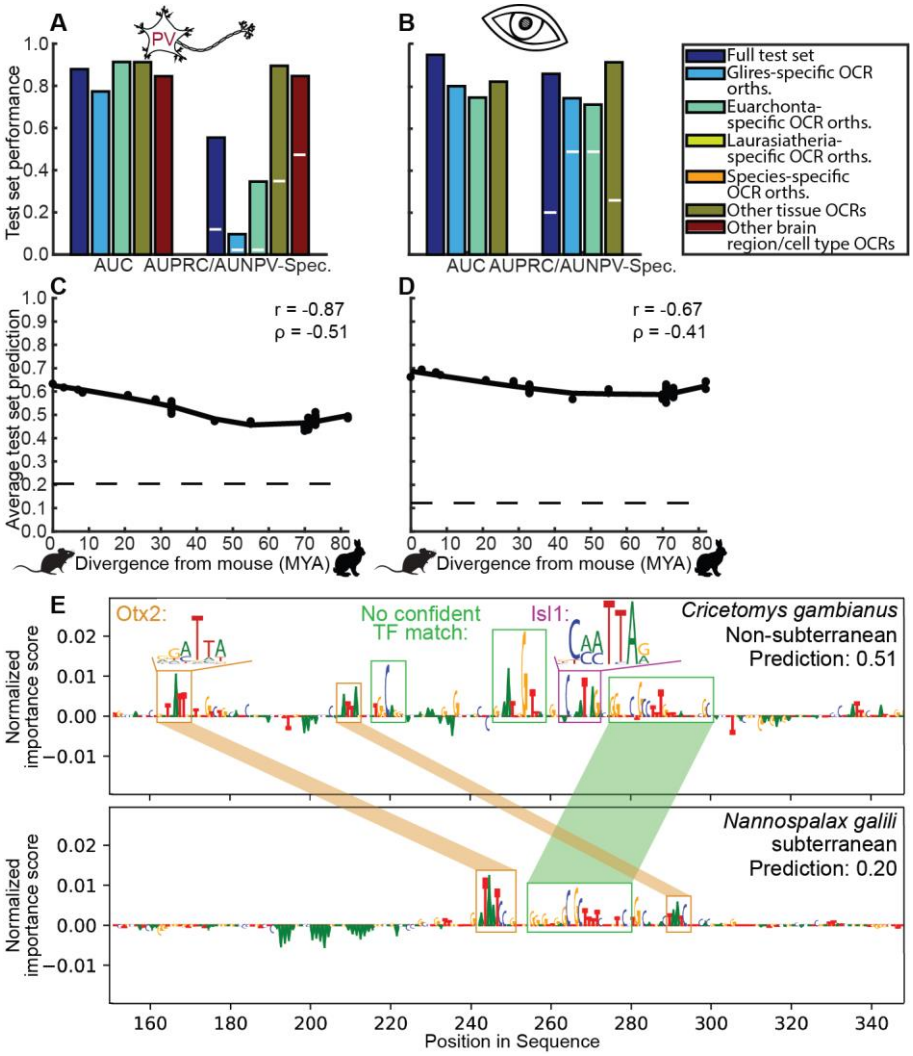

Figure 3: PV+ interneuron and retina multi-species CNN performance. (**A-B**) show the AUROC and the AUPRC/NPV-Spec. for the full test set, clade-specific OCRs and non-OCRs, and shared versus tissue/cell type-specific OCRs and non-OCRs for multi-species PV+ interneuron (**A**) and retina (**B**) CNNs. (**C-D**) show the negative relationship between the average house mouse OCR ortholog multi-species PV+ interneuron (**C**) and retina (**D**) CNN predictions for Glires species and the MYA when each species diverged from house mouse. E shows the multi-species retina model normalized importance scores for each position in the summit +/- 100bp of an OCR near *OTX2* that was previously shown to have a higher relative evolutionary rate in subterranean mammals. Orange boxes mark matches to the house mouse Otx2 motif, the magenta box marks the match to the house mouse Isl1 motif, and green boxes mark regions with high importance scores that do not match any known TF motif. Motifs were downloaded from CIS-BP (*86*) and visualized using meme2images from the MEME suite (*87*). No nucleotides in either ortholog outside these central 200 base pairs had a normalized importance score with absolute value greater than one.

mammals (*9*). This OCR's ortholog in *Nannospalax galili*, a subterranean mole-rat, was confidently predicted to be closed, while its ortholog in a non-subterranean pouched rat, *Cricetomys gambianus*, the most closely related mammal in Zoonomia that never lives underground (diverged ~45 MYA (*36*)), was

predicted to be open. Both of these OCR orthologs contained two motifs for Otx2 as well as a third motif that could not be easily interpreted with high importance scores. In addition to those important sequences, the *Cricetomys gambianus* ortholog had a high importance score for the motif for Isl1, a transcription factor involved in the development of bipolar and cholinergic amacrine cells of the retina (*37*). There were also two additional sequences with high importance scores unique to *Cricetomys gambianus* relative to *Nannospalax galili* that did not match any known TF motif, demonstrating the value of using a modeling strategy that does not require featurizing the sequence based on known information (**Fig. 3E**).

From the four cross-species OCR datasets of interest (motor cortex, liver, PV+ interneuron, and retina), we identify 50,942,699 total orthologous regions across 222 Boreoeutherian mammals from 402,880 total OCRs. Relative to human OCR annotations and phyloP annotations alone, we find that these predictions can provide a substantial boost for interpreting human disease-associated loci, with greater tissue- and cell type specificity. For example, in our other work, we found that human orthologs of regions predicted to have conserved motor cortex open chromatin are enriched for overlapping SNPs associated with schizophrenia, while human orthologs of regions predicted to have conserved liver open chromatin are enriched for overlapping SNPs associated with cholesterol-related traits (*38*, *39*). These results demonstrate the power of TACIT to identify functionally relevant patterns of conservation.

Applying TACIT to mammalian phenotypes

A framework for associating predicted open chromatin with phenotypes

Having trained models to predict open chromatin status of OCR orthologs in four tissues and cell types – motor cortex, liver, retina, and PV+ interneurons within the motor cortex – we identified individual OCRs whose predicted open chromatin across species is associated with phenotypes (**Fig. 1**). We applied the phylolm and phyloglm methods (*15*) for continuous and binary traits, respectively. These methods are modifications of phylogenetic generalized least squares (*40*, *41*) designed for faster performance. We

used them to test for a relationship between one OCR ortholog's open chromatin predictions across species and phenotype annotations across species that cannot be explained by the species phylogeny alone. To minimize false positives, we implemented phylogenetic permulations (*16*), enabling us to evaluate the significance of each OCR-phenotype relationship against a background distribution of shuffled phenotypes with similar phylogenetic structures (**Materials and Methods**).

TACIT identifies motor cortex and PV+ interneuron OCRs associated with the evolution of brain size

We used TACIT to identify motor cortex OCRs whose predicted open chromatin across mammals is significantly associated with brain size, a complex trait with great diversity across mammals that is thought to underlie human cognitive ability (*42*). As brain size scales with body size, we used the brain size residual (brain mass minus the predicted value of brain mass from a regression on body mass), which we obtained for 158 boreoeutherian mammals (*43, 44*). Before applying TACIT, we investigated whether there are proteins whose relative evolutionary rates (*19*) are associated with the evolution of brain size residual. We did not find any proteins with a significant association after RERconverge's default multiple hypothesis correction (corrected $p \geq 0.05$ for all genes) (*19, 45*), which corroborates evidence that the top decile of TFs with the highest fraction of conserved base pairs tend to be enriched for embryonic development and brain function (PhyloP ≥ 2.241, FDR $< 5\%$) (*39*) and previous work suggesting that enhancer loss drove the evolution of human-specific patterns in brain growth (*10*). In contrast, using TACIT, we found 34 motor cortex OCRs with a significant association with brain size residual after false discovery rate correction (α=0.05). We then examined all genes near (TSSs within 1Mb) those OCRs. Of the associated OCRs, 29 are near genes whose corresponding proteins play important roles in brain development, and 6 are near genes whose corresponding proteins are involved in brain tumor growth (**Table S8**). While many of these genes may influence brain size during development, the OCRs that regulate them might continue to be open during adulthood. This would be consistent with recent evidence that neural progenitors are responsible for the evolution of brain size in the great apes (*46*).

Of the 29 brain size residual-associated OCRs near brain development genes, 23 are near genes with mutations that cause neurological disorders, including 8 OCRs near genes in which mutations have been reported to cause microcephaly or macrocephaly (**Table S8, Figs. S8A-H**) (*47*). Furthermore, we found that the p-values of all motor cortex OCRs whose human orthologs are near (in hg38 coordinates) genes mutated in microcephaly or macrocephaly have a significantly lower distribution than the p-values of other motor cortex OCRs with human orthologs (p=0.0073, 1-sided Wilcoxon rank-sum test).

We identified two OCRs near *SATB1* — a gene with both microcephaly- and macrocephaly-associated mutations (*48*) — whose motor cortex predicted open chromatin status is significantly associated with brain size residual (**Fig. 4A-B, Figs. S8D**,**H**). For both of these associations, predicted open chromatin is associated with small brain size residual. The OCRs' coordinates in the genomes in which they were initially identified are chr17:52351209-52351928 (mm10) and chr2:174466184-174466517 (rheMac8). They are each about 500kb from the TSS of the gene, where one is upstream and the other is downstream. Neither OCR is near any other gene with a known connection to brain development; *Satb1*/*SATB1* is the second-closest gene to each, and the closer genes, *Kcnh8* and *TBC1D5*, each have known roles outside of brain growth (*49*, *50*). The associations seem to be driven in large part by, respectively, cetaceans (**Fig. 4A**) and great apes (**Fig. 4B**), both of which have a large variation in brain size (*51*). In particular, the latter OCR is predicted to be active in all great apes except for humans, the great ape with the largest brain size residual. Interestingly, the reported case of *SATB1*-associated macrocephaly at birth was caused by a mutation that disrupts a large portion of the protein product, while microcephaly was usually reported with *SATB1* missense mutations (*48*). This pattern is consistent with the significant negative associations between predicted open chromatin and brain size residual, assuming that the OCRs we identified positively regulate the expression of *SATB1*.

We identified another OCR, chr2:75345159-75346046 (rheMac8), whose predicted motor cortex open chromatin also has a strong negative association with brain size residual in cetaceans (**Fig. 4C**). The closest gene to this OCR is *LRIG1*, which is about 250kb from the OCR. *LRIG1* slows and delays the

237 differentiation of neural stem cells (*52*, *53*). While this OCR is also near other genes, none of those genes

238 have a known role in brain size.

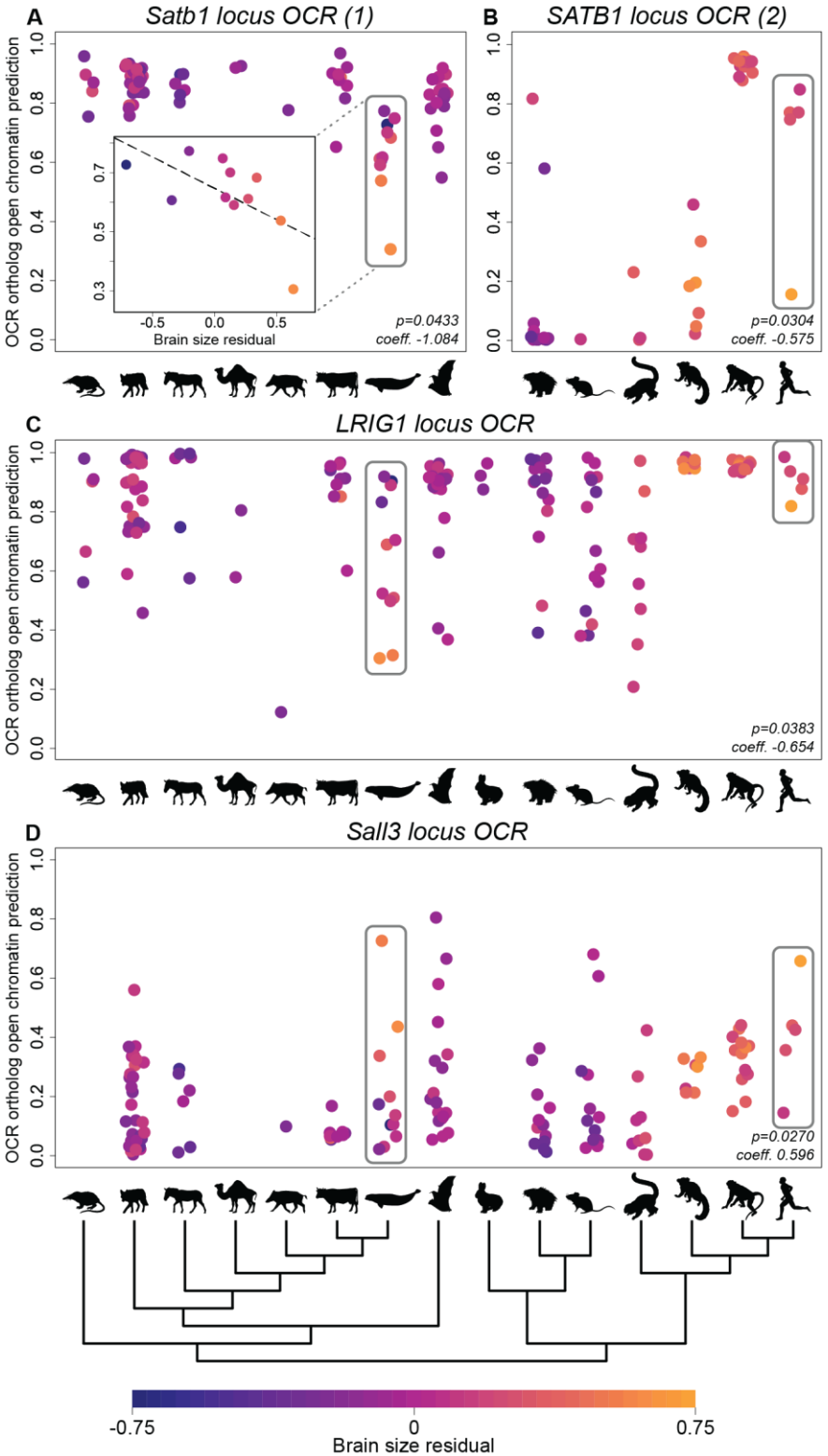

Figure 4: Examples of associations between predicted motor cortex OCR ortholog open chromatin and brain size residual. (**A**-**B**) highlight the negative association between predicted motor cortex open chromatin and brain size residual of orthologs of two motor cortex OCRs in the *SATB1* locus, chr17:52351209-52351928 (mm10) and chr2:174466184-174466517 (rheMac8), within Laurasiatheria and Euarchontoglires, respectively. The latter OCR has no orthologs in Lagomorpha, which is omitted from panel (**B**). Boreoeutherian mammal-wide panels are shown in Fig. S9. (**C**) highlights the negative association of orthologs of a motor cortex OCR in the *LRIG1* locus, chr15:40082805-40083380 (mm10). (**D**) highlights the positive association of orthologs of a motor cortex OCR in the *Sall3* locus, chr18:81802310-81802951 (mm10). Each point represents one ortholog; they are grouped along the x-axis of each panel by clade as shown by the tree below. The clades and example species are listed in Table S10. The hominoid and cetacean clades are highlighted by gray boxes in each panel. Points are colored by brain size residual following the scale at the bottom. The permulations p-value after Benjamini-Hotchberg correction and the coefficient on the predicted open chromatin returned by phylolm are shown in the lower right of each panel.

Also among the OCRs we identified near brain development genes is an OCR, chr18:81802310-81802951 (mm10), about 800kb from the gene *Sall3*. Sall3 is the fourth-closest gene to this OCR, and one closer gene, *Mbp*, does have a connection to brain development (*54*). Hi-C from adult human cortex (*55*) shows that the bin containing the human ortholog of this OCR is close to *SALL3* in 3D space ($p=2.3 \times 10^{-11}$, **Table S8**) but is not close to MBP ($p=1$). This OCR displays a positive association with brain size residual both overall and within mammalian clades with especially large variations in brain size, including the great apes and cetaceans (**Fig. 4D**). *Sall3* is a member of the spalt-like family of transcription factors, which are important in development (*56*). Although a specific role of *Sall3* in motor cortex has not been described, there is evidence that *Sall3* regulates the maturation of neurons in other regions of the brain (*57*, *58*), and *Sall3* is expressed in developing motor neurons (*58*) and human cerebral cortex (*59*).

We extended our framework to establish Cell-TACIT, a version of TACIT that identifies OCRs in specific cell types (*60, 61*) whose open chromatin predictions are associated with a phenotype of interest. We used Cell-TACIT for PV+ interneurons within the motor cortex to identify such OCRs whose predicted activity across Euarchontoglires is significantly associated with brain size residual. PV+ interneurons are a minority population, representing roughly 4 - 8% of neurons and 2 - 4 % of the total cell population in the mouse cortex (*62*) yet are critical in cortical microcircuits and human brain disorders like schizophrenia (*63, 64*). Given this sparsity, our bulk motor cortex open chromatin data may not capture OCRs that are specific to PV+ interneurons. In fact, about 30% of mouse PV+ OCRs do not overlap any bulk motor cortex OCRs, including non-reproducible peaks. We identified 13 OCRs whose predicted open chromatin in PV+ interneurons is associated with species' brain size residuals after false discovery rate correction ($\alpha=0.05$) (**Table S9**), 11 of which are house mouse OCRs for which predicted open chromatin is associated with having a smaller brain size residual.

We identified three PV+ interneuron OCRs that are significantly negatively associated with brain size residual and are within 1Mb of a gene that is mutated in macrocephaly or microcephaly (**Table S9**, **Figs. S8I-K**). Two of those OCRs — chr13:114757413-114757913 (mm10) and chr13:114793237-

114793737 (mm10) — are respectively about 60kb and 25kb from the *Mocs2* gene. Both have strong associations with brain size residual within Euarchonta (primates and their closest relatives), especially Hominoidea, and the first also has some association within Glires (rodents and their closest relatives) (**Fig. 5A-B**, respectively). *Mocs2* is one of four genes involved in Molybdenum cofactor biosynthesis (*65*). Molybdenum cofactor deficiency (MoCD) in humans is a rare, fatal disease marked by intractable seizures, hypoxia, and microcephaly (*66*). We also identified an OCR, chr1:95762160-95762660 (mm10), that is about 100kb away from the gene *St8sia4*, which is important for the development and density of interneurons — including PV+ interneurons — in the cortex (*67*, *68*).

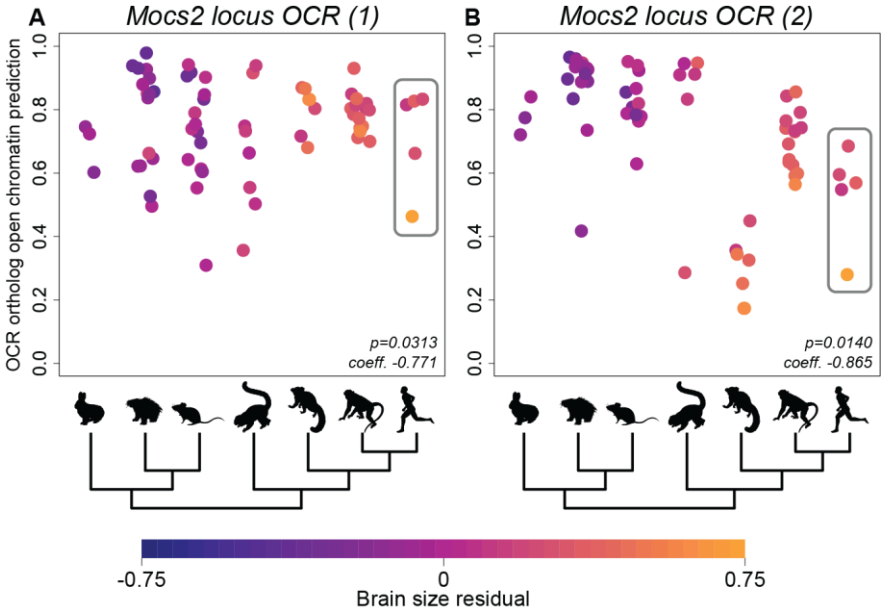

Figure 5: Examples of associations between predicted PV+ interneuron OCR ortholog open chromatin and brain size residual.
(**A-B**) highlight the negative association within Euarchontoglires between predicted PV+ interneuron open chromatin and brain size residual of orthologs of two PV+ interneuron OCRs in the *Mocs2* locus, chr13:114757413-114757913 (mm10) and chr13:114793237-114793737 (mm10). Each point represents one ortholog; they are grouped along the x-axis of each panel by clade as shown by the tree below. The clades and example species are listed in Table S10. The hominoid clade is highlighted by a gray box in each panel. Points are colored by brain size residual following the scale at the bottom.

Interestingly, there is no overlap between the bulk motor cortex OCRs and PV+ interneuron OCRs with predicted activity that is significantly associated with brain size residual. In fact, no house mouse OCR ortholog from either set is within 5Mb of a house mouse OCR ortholog from the other set. We also investigated liver OCRs associated with brain size residual and found that none of these OCRs

277 overlapped the associated motor cortex OCRs (**Supplementary Text**) (*26*). This highlights the
278 complementary information provided by using TACIT OCRs from different tissues as well as from using
279 both TACIT and Cell-TACIT.
280
281 *Cell-TACIT and TACIT identify PV+ interneuron and motor cortex open chromatin regions in loci*
282 *associated with the evolution of social living*
283
284 One challenge of using TACIT and Cell-TACIT is that tens to hundreds of thousands of OCRs are tested,
285 which requires correcting for large numbers of hypotheses. This is necessary for applying TACIT to
286 phenotypes like brain size for which there is no strict subset of the genome that is known to be involved
287 in the phenotype. In contrast, when such a subset is known, we can increase power by restricting OCRs to
288 those in that subset. We used this targeted approach to examine relationships between solitary and group
289 living lifestyles and predicted PV+ OCR activity within the 1,661,222bp Williams-Beuren Syndrome
290 (WBS) deletion region (**Fig. 6A**), where haploinsufficiency causes increased sociability, intellectual
291 disability, and enhanced verbal fluency in human patients (*69*). Although the WBS locus has not been
292 linked to PV+ interneurons specifically, PV+ interneurons are well-known for their involvement in social
293 behaviors and neuropsychiatric disorders with social components such as autism spectrum disorder
294 (ASD) and schizophrenia (*70*). Molecular evidence for PV+ interneuron involvement suggests associated
295 transcriptional changes. For example, *PVALB* was the most strongly downregulated transcript in ASD
296 brain tissue compared to healthy controls and in animal models of monogenetic neurodevelopmental
297 syndromic disorders (*71*, *72*), and single-nucleus RNA-seq from schizophrenia brain tissue revealed more
298 abnormal gene expression in PV+ interneurons than in any other neuronal cell type (*73*, *74*). Direct
299 expression manipulation of psychiatric genes in PV+ interneurons was shown to induce social deficits in
300 mice, whereas similar manipulations in other neuron cell types had different effects (*75*).
301 The Mesozoic ancestors of today's mammals were likely primarily solitary-living, defined by
302 separate foraging and home ranges for females (*76*). Following the End-Cretaceous Mass Extinction,

many extant lineages in disparate clades evolved towards social living strategies, including group living and breeding pair monogamy (*76*). Given the impact of PV+ neuron gene expression on social behaviors, we hypothesized that there might be PV+ OCR evolution associated with social structure transitions in mammals.

Before investigating our results, we evaluated whether Cell-TACIT was producing reliable results by comparing results from Cell-TACIT run genome-wide on PV+ OCR orthologs to locations of human genome-wide association study (GWAS) hits for schizophrenia, a disorder associated with solitariness. Specifically, we divided PV+ OCRs into two groups: those that overlapped a schizophrenia-associated variant and those that did not (*77*). We determined the strength of association of all OCRs with solitary living in mammals. The set of PV+ interneuron OCRs with schizophrenia-associated variants had a shifted phyloglm p-value distribution for association with solitary living compared to the p-value distribution for other PV+ interneuron OCRs (one-sided Wilcoxon rank-sum $p = 0.035$) (**Fig. 6B**).

When applying Cell-TACIT to only the WBS locus, we identified a mouse OCR (out of two OCRs in this locus) 29kb upstream of *GTF2IRD1* (human ortholog is 36kb upstream) that was marginally associated with non-solitary living ($p = 0.08$) (**Fig. 6C**) and associated with group living ($p = 0.02$). To evaluate whether this association was limited to PV+ interneurons, we also evaluated the relationship between predicted bulk motor cortex open chromatin and solitary as well as group living. For solitariness, we found one significantly negatively associated OCR ($p = 0.005$) (**Fig. 6D**). This OCR is in an intron of *GTF2IRD1* that is about 26kb from its nearest TSS but does not overlap the OCR identified for PV+ interneurons. For group living, we found two significantly associated OCRs, one of which is negatively associated ($p = 0.04$) and the other of which is positively associated ($p = 0.008$) and is the same OCR we found for solitariness. Of the 27 protein-coding genes in the WBS locus, *GTF2IRD1* is one of only two genes, where the other gene is its neighbor (*GTF2I*), with structural variants associated with the extreme sociability in dogs that makes them easier to domesticate than wolves (*78*). We additionally evaluated the

327 relationship between predicted liver open chromatin and solitary as well as group living but did not obtain

328 any statistically significant relationships after multiple hypothesis correction.

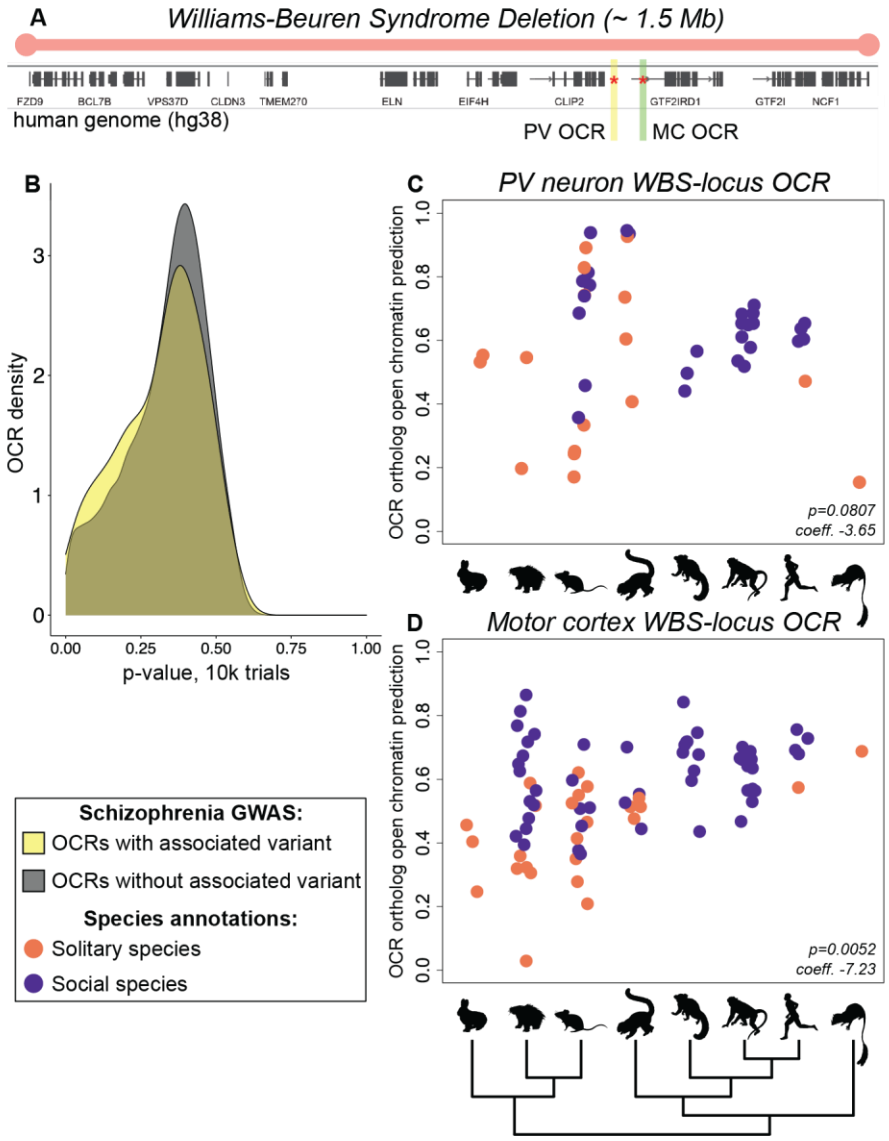

Figure 6: Associations between predicted PV+ interneuron and motor cortex OCR ortholog open chromatin and solitary living.
(**A**) A visualization of the human WBS deletion region. The locations of the PV+ interneuron and motor cortex OCRs (highlighted in panels (**C**) and (**D**)) near the gene *GTF2IRD1* are shown in yellow and green, respectively. (**B**) shows the difference in p-value distributions for association between solitary living and predicted open chromatin of PV+ interneuron OCRs whose human ortholog overlaps schizophrenia GWAS SNPs versus all other PV+ interneuron OCRs with a human ortholog. (**C**) highlights the marginal negative association between predicted PV+ interneuron open chromatin and solitary living of orthologs of a PV+ interneuron OCR near *GTF2IRD1*, chr5:134485808-134486308 (mm10). (**D**) highlights the negative association between predicted motor cortex open chromatin and solitary living of orthologs of a motor cortex OCR near *GTF2IRD1*, chr3:42408296-42408946 (rheMac8). For panels (**C-D**), each point represents one ortholog; they are grouped along the x-axis of each panel by clade as shown by the tree below. The clades and example species are listed in Table S10. Points are colored to indicate solitary versus social living following the key at the lower left.

TACIT and Cell-TACIT identify open chromatin regions associated with the evolution of vocal learning

We applied TACIT and Cell-TACIT to vocal learning, the ability to modify vocal output as a result of social experience, which has convergently evolved across mammals and been associated with convergent patterns of gene expression in the motor cortex (*2*, *79*, *80*). We identified 42 OCRs displaying convergent patterns of predicted open chromatin after false discovery rate correction (α=0.05) for motor cortex tissue and 14 for PV+ interneurons, which are described in more depth in our other work . Notably, these vocal learning-associated OCRs showed some concordance with results obtained using complementary methods for detecting convergent evolution. One of the motor cortex OCRs lies 88kb from *Vip*, whose expression in the motor cortex has been associated with vocal learning (*2*). Another OCR is 715kb from *TSHZ3*, whose amino acid sequence also showed convergent evolution associated with vocal learning behavior ($p < 0.0001$, rank 3) (*81*). *TSHZ3* is involved in the formation of cortico-striatal circuits, which play a central role in vocal learning behavior in mammals and birds, and its disruption in the human population is associated with a form of autism that includes delayed or disrupted speech acquisition (*80*, *82*).

DISCUSSION

We present TACIT and Cell-TACIT, new methods for associating genotypes to phenotypes based on machine learning predictions of tissue- or cell type-specific open chromatin. Our approach overcomes the limitations of nucleotide-level conservation-based approaches, which cannot completely account for the conservation of enhancer function in the presence of low sequence conservation and cannot capture the tissue- and cell type-specificity of enhancer activity (*25*), because our machine learning models learn the conserved regulatory code underlying enhancer activity in our tissue or cell type of interest. We provide a community resource of annotated predicted open chromatin for more than 400,000 OCRs from four tissues and cell types across 222 mammalian species.

We applied TACIT and Cell-TACIT to identify tissue- and cell type-specific OCRs whose predicted open chromatin status across species is associated with brain size residual, solitary living, group living, and vocal learning, including OCRs near genes that were previously implicated in these phenotypes. Specifically, we identified motor cortex and PV+ interneuron OCRs associated with brain size residual that are near genes whose mutations are associated with microcephaly and macrocephaly, as well as motor cortex OCRs with a strong brain size residual association in Cetaceans, which provide candidate mechanisms for the evolution of brain size beyond the previously identified human-specific deletion (*10*). In addition, the WBS deletion region OCRs with the strongest evolution of solitary and group living association are near a critical gene for WBS presentation as well as canine social behavior (*78*). Genome-wide, the associations of PV+ interneuron OCRs with group and solitary living are correlated with whether the OCR overlaps a GWAS hit for schizophrenia, which suggests that OCRs involved in the evolution of traits may also be involved in disorders associated with those traits, a result further supported by our other work (*38*). To be confident that the OCRs we identified have enhancer activity that differs between species, we would need to use reporter assays to test the OCR orthologs' enhancer activity in multiple species. In addition, to thoroughly demonstrate that these OCRs regulate the nearby genes associated with the phenotypes, we would need to do experiments like CRISPR followed by RNA-qPCR to knock out the OCR and show that the knock-out causes a change in the expression of the nearby gene. Furthermore, considering genes with TSSs within 1Mb may limit our ability to identify real gene-OCR relationships (*83*), but, as data measuring three-dimensional genome interactions becomes available at higher resolution and in additional species, tissues, and cell types, our ability to link candidate enhancers associated with phenotypes to the genes they likely regulate will improve.

While our previous work used data from three species for model-training (*25*), in this work, we developed a new strategy for negative set construction that allowed us to train accurate models using data from only two species. Our success in doing this enabled us to train models that accurately predict whether sequence differences across species in PV+ interneuron OCR orthologs are associated with PV+ interneuron open chromatin changes, demonstrating that the regulatory code is conserved across

Euarchontoglires not only at the bulk tissue level but also in a specific neuronal cell type. We also found that having data from more clades enabled us to identify OCRs associated with phenotypes in additional clades, such as the OCR near *LRIG1* associated with the evolution of brain size residual in the Cetacea order within Laurasiatheria, and provides us with the power to identify OCRs with weaker associations with a phenotype across multiple lineages, such as the OCR near *Sall3* associated with the evolution of brain size residual in both Euarchonta and Laurasiatheria.

Unlike phyloP or PhastCons scores, the broad application of TACIT and Cell-TACT is limited by the availability of high-quality open chromatin data from the same tissue or cell type in multiple species. TACIT and Cell-TACIT require enhancer activity data from at least two species for evaluating machine learning models, and, to limit confounding factors, the data should ideally contain animals at comparable developmental stages, biological replicates from both sexes, and animals that were sacrificed in comparable behavioral states at approximately the same relative time in their circadian cycles. Additionally, predictions are currently limited to orthologs of experimentally identified candidate enhancers, meaning that we are not able to capture enhancers that are not active in the experimentally assayed species, cell types, developmental stages, or conditions. Furthermore, our approach assumes that the evolution of a phenotype is controlled by the same candidate enhancer across species, but there are likely many phenotypes controlled by genes that are not activated by the same enhancer in every species. We also treat missing or unusable OCR orthologs as missing data, but some of these are likely lost OCRs. Exciting extensions to our approach include training models to accurately predict whether sequence differences cause changes in candidate enhancer activity genome-wide, jointly modeling cross-species predicted enhancer activity of enhancers near the same gene, and using genome quality and the predicted open chromatin of OCRs in closely related species to determine when a lack of a usable OCR ortholog should be treated as a negative. Finally, our approach assumes that the regulatory code in our tissue or cell type of interest is conserved across the species we are testing, an assumption that may be violated in some tissues and cell types. For example, this may explain the sub-optimal performance of our retina CNNs trained on mouse sequences in predicting Euarchonta-specific open and closed chromatin (*84*, *85*).

With the Zoonomia Cactus alignment of over two hundred mammalian genomes and the wealth of publicly available enhancer activity data from matching tissues and cell types in human, house mouse, and some other species, TACIT and Cell-TACIT can currently be applied to identify candidate enhancers associated with the evolution of many mammalian phenotypes. Because TACIT and Cell-TACIT require enhancer activity data from tissues or cell types of interest in only a few species, they can be used to identify losses of enhancer activity associated with changes in a phenotype in challenging-to-study species for which we have genomes but cannot collect tissue samples. In addition, while we trained our models for TACIT using open chromatin, TACIT can also be applied using other assays of enhancer activity, such as H3K27ac and EP300 ChIP-seq (*27*). Candidate enhancers associated with the evolution of phenotypes near genes involved in diseases related to those phenotypes may provide insights into disease mechanisms. We anticipate that, as more genomes and regulatory genomics data become available, TACIT and Cell-TACIT will provide insights into the regulatory mechanisms governing a wide range of phenotypes.

LIST OF SUPPLEMENTARY MATERIALS

Materials and Methods

Supplementary Text

Supplementary Figures 1-8

Supplementary Tables 1-11

REFERENCES AND NOTES

1. M. King, A. Wilson, Evolution at two levels in humans and chimpanzees. *Science*. **188**, 107–116

(1975).

2. 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. 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).

3. A. A. Fushan, A. A. Turanov, S. G. Lee, E. B. Kim, A. V. Lobanov, S. H. Yim, R. Buffenstein, S. R. Lee, K. T. Chang, H. Rhee, J. S. Kim, K. S. Yang, V. N. Gladyshev, Gene expression defines natural changes in mammalian lifespan. *Aging Cell*. **14**, 352–365 (2015).

4. G. A. Wray, The evolutionary significance of cis-regulatory mutations. *Nat. Rev. Genet.* **8**, 206–216 (2007).

5. D. Villar, P. Flicek, D. T. Odom, Evolution of transcription factor binding in metazoans — mechanisms and functional implications. *Nat. Rev. Genet.* **15**, 221–233 (2014).

6. E. Z. Kvon, O. K. Kamneva, 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).

7. L. A. Lettice, A. E. Hill, P. S. Devenney, R. E. Hill, Point mutations in a distant sonic hedgehog cis-regulator generate a variable regulatory output responsible for preaxial polydactyly. *Hum. Mol. Genet.* **17**, 978–985 (2008).

8. D. Furniss, L. A. Lettice, I. B. Taylor, P. S. Critchley, H. Giele, R. E. Hill, A. O. M. Wilkie, A variant in the sonic hedgehog regulatory sequence (ZRS) is associated with triphalangeal thumb and deregulates expression in the developing limb. *Hum. Mol. Genet.* **17**, 2417–2423 (2008).

9. 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).

10. C. Y. McLean, P. L. Reno, A. A. Pollen, A. I. Bassan, T. D. Capellini, C. Guenther, V. B. Indjeian, X. Lim, D. B. Menke, B. T. Schaar, A. M. Wenger, G. Bejerano, D. M. Kingsley, Human-specific loss of regulatory DNA and the evolution of human-specific traits. *Nature*. **471**, 216–219 (2011).

11. Zoonomia Consortium, A comparative genomics multitool for scientific discovery and conservation. *Nature*. **587**, 240–245 (2020).

12. J. Armstrong, G. Hickey, M. Diekhans, I. T. Fiddes, A. M. Novak, A. Deran, Q. Fang, D. Xie, S. Feng, J. Stiller, 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).

13. G. Hickey, B. Paten, D. Earl, D. Zerbino, D. Haussler, HAL: a hierarchical format for storing and analyzing multiple genome alignments. *Bioinformatics*. **29**, 1341–1342 (2013).

14. X. Zhang, I. M. Kaplow, M. Wirthlin, T. Y. Park, A. R. Pfenning, HALPER facilitates the identification of regulatory element orthologs across species. *Bioinformatics*. **36**, 4339–4340 (2020).

15. L. S. T. Ho, C. Ané, A linear-time algorithm for Gaussian and non-Gaussian trait evolution models. *Syst. Biol.* **63**, 397–408 (2014).

16. E. Saputra, A. Kowalczyk, L. Cusick, N. Clark, M. Chikina, Phylogenetic Permulations: A Statistically Rigorous Approach to Measure Confidence in Associations in a Phylogenetic Context. *Mol. Biol. Evol.* **38**, 3004–3021 (2021).

17. 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).

18. Z. Yang, PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**, 1586–1591 (2007).

19. 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).

20. J. G. Roscito, K. Sameith, G. Parra, B. E. Langer, A. Petzold, C. Moebius, M. Bickle, M. T. Rodrigues, M. Hiller, Phenotype loss is associated with widespread divergence of the gene regulatory landscape in evolution. *Nat. Commun.* **9**, 4737 (2018).

21. D. Villar, C. Berthelot, S. Aldridge, T. F. Rayner, M. Lukk, M. Pignatelli, T. J. Park, R. Deaville, J. T. Erichsen, A. J. Jasinska, J. M. A. Turner, M. F. Bertelsen, E. P. Murchison, P. Flicek, D. T. Odom, Enhancer Evolution across 20 Mammalian Species. *Cell*. **160**, 554–566 (2015).

22. K. Lindblad-Toh, M. Garber, O. Zuk, M. F. Lin, B. J. Parker, S. Washietl, P. Kheradpour, J. Ernst, G. Jordan, E. Mauceli, L. D. Ward, C. B. Lowe, A. K. Holloway, M. Clamp, S. Gnerre, J. Alföldi, K. Beal, J. Chang, H. Clawson, J. Cuff, F. Di Palma, S. Fitzgerald, P. Flicek, M. Guttman, M. J. Hubisz, D. B. Jaffe, I. Jungreis, W. J. Kent, D. Kostka, M. Lara, A. L. Martins, T. Massingham, I. Moltke, B. J. Raney, M. D. Rasmussen, J. Robinson, A. Stark, A. J. Vilella, J. Wen, X. Xie, M. C. Zody, Broad Institute Sequencing Platform and Whole Genome Assembly Team, J. Baldwin, T. Bloom, C. W. Chin, D. Heiman, R. Nicol, C. Nusbaum, S. Young, J. Wilkinson, K. C. Worley, C. L. Kovar, D. M. Muzny, R. A. Gibbs, Baylor College of Medicine Human Genome Sequencing Center Sequencing Team, A. Cree, H. H. Dihn, G. Fowler, S. Jhangiani, V. Joshi, S. Lee, L. R. Lewis, L. V. Nazareth, G. Okwuonu, J. Santibanez, W. C. Warren, E. R. Mardis, G. M. Weinstock, R. K. Wilson, Genome Institute at Washington University, K. Delehaunty, D. Dooling, C. Fronik, L. Fulton, B. Fulton, T.

Graves, P. Minx, E. Sodergren, E. Birney, E. H. Margulies, J. Herrero, E. D. Green, D. Haussler, A. Siepel, N. Goldman, K. S. Pollard, J. S. Pedersen, E. S. Lander, M. Kellis, A high-resolution map of human evolutionary constraint using 29 mammals. *Nature*. **478**, 476–482 (2011).

23. V. Snetkova, A. R. Ypsilanti, J. A. Akiyama, B. J. Mannion, I. Plajzer-Frick, C. S. Novak, A. N. Harrington, Q. T. Pham, M. Kato, Y. Zhu, J. Godoy, E. Meky, R. D. Hunter, M. Shi, E. Z. Kvon, V. Afzal, S. Tran, J. L. R. Rubenstein, A. Visel, L. A. Pennacchio, D. E. Dickel, Ultraconserved enhancer function does not require perfect sequence conservation. *Nat. Genet.* **53**, 521–528 (2021).

24. 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. McGlinn, M. Francois, B. M. Degnan, Deep conservation of the enhancer regulatory code in animals. *Science*. **370**, eaax8137 (2020).

25. I. M. Kaplow, D. E. Schäffer, M. E. Wirthlin, A. J. Lawler, A. R. Brown, M. Kleyman, A. R. Pfenning, https://www.biorxiv.org/content/10.1101/2020.12.04.410795v3.

26. Materials and methods are available as supplementary materials at the Science website.

27. L. Chen, A. E. Fish, J. A. Capra, Prediction of gene regulatory enhancers across species reveals evolutionarily conserved sequence properties. *PLoS Comput. Biol.* **14**, e1006484 (2018).

28. D. R. Kelley, Cross-species regulatory sequence activity prediction. *PLoS Comput. Biol.* **16**, e1008050 (2020).

29. L. Minnoye, I. I. Taskiran, D. Mauduit, M. Fazio, L. Van Aerschot, G. Hulselmans, V. Christiaens, S. Makhzami, M. Seltenhammer, P. Karras, A. Primot, E. Cadieu, E. van Rooijen, J.-C. Marine, G. Egidy, G.-E. Ghanem, L. Zon, J. Wouters, S. Aerts, Cross-species analysis of enhancer logic using deep learning. *Genome Res.* **30**, 1815–1834 (2020).

30. A. Shrikumar, P. Greenside, A. Kundaje, Learning important features through propagating activation

519 differences. *Proceedings of the 34th International Conference on Machine Learning*. **70**, 3145–3153
520 (2017).

521 31. S. M. Lundberg, S.-I. Lee, A Unified Approach to Interpreting Model Predictions. *Advances in*
522 *Neural Information Processing Systems*. **31**, 4768–4777 (2017)

523 32. A. J. Harrington, A. Raissi, K. Rajkovich, S. Berto, J. Kumar, G. Molinaro, J. Raduazzo, Y. Guo, K.
524 Loerwald, G. Konopka, K. M. Huber, C. W. Cowan, MEF2C regulates cortical inhibitory and
525 excitatory synapses and behaviors relevant to neurodevelopmental disorders. *Elife*. **5**, e20059
526 (2016).

527 33. Y. C. Chen, H. Y. Kuo, U. Bornschein, H. Takahashi, S. Y. Chen, K. M. Lu, H. Y. Yang, G. M.
528 Chen, J. R. Lin, Y. H. Lee, Y. C. Chou, S. J. Cheng, C. T. Chien, W. Enard, W. Hevers, S. Pääbo, A.
529 M. Graybiel, F. C. Liu, Foxp2 controls synaptic wiring of corticostriatal circuits and vocal
530 communication by opposing Mef2c. *Nat. Neurosci.* **19**, 1513–1522 (2016).

531 34. J. P. Babeu, F. Boudreau, Hepatocyte nuclear factor 4-alpha involvement in liver and intestinal
532 inflammatory networks. *World J. Gastroenterol.* **20**, 22–30 (2014).

533 35. D. Alpern, D. Langer, B. Ballester, S. Le Gras, C. Romier, G. Mengus, I. Davidson, TAF4, a subunit
534 of transcription factor II D, directs promoter occupancy of nuclear receptor HNF4A during post-natal
535 hepatocyte differentiation. *Elife*. **3**, e03613 (2014).

536 36. S. Kumar, G. Stecher, M. Suleski, S. B. Hedges, TimeTree: A Resource for Timelines, Timetrees,
537 and Divergence Times. *Mol. Biol. Evol.* **34**, 1812–1819 (2017).

538 37. Y. Elshatory, D. Everhart, M. Deng, X. Xie, R. B. Barlow, L. Gan, Islet-1 controls the differentiation
539 of retinal bipolar and cholinergic amacrine cells. *J. Neurosci.* **27**, 12707–12720 (2007).

540 38. B. N. Phan, A. J. Lawler, J. He, I. M. Kaplow, A. R. Brown, D. E. Schäffer, W. R. Stauffer,

Zoonomia Project Consortium, A. R. Pfenning, Fine-mapping candidate neurological disease-associated regulatory variants using cell type-aware comparative genomics. *Submitted to Science with Zoonomia Consoritum*.

39. P. Sullivan, Zoonomia Consortium, Evolutionary constraint uniquely informs human disease mechanism. *Submitted to Science with Zoonomia Consoritum*.

40. A. Grafen, The phylogenetic regression. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **326**, 119–157 (1989).

41. A. R. Ives, T. Garland Jr, Phylogenetic logistic regression for binary dependent variables. *Syst. Biol.* **59**, 9–26 (2010).

42. C. Mitchell, D. L. Silver, Enhancing our brains: Genomic mechanisms underlying cortical evolution. *Semin. Cell Dev. Biol.* **76**, 23–32 (2018).

43. J. R. Burger, M. A. George, C. Leadbetter, F. Shaikh, The allometry of brain size in mammals. *J. Mammal.* **100**, 276–283 (2019).

44. S. Herculano-Houzel, The remarkable, yet not extraordinary, human brain as a scaled-up primate brain and its associated cost. *Proc. Natl. Acad. Sci. U. S. A.* **109 Suppl 1**, 10661–10668 (2012).

45. Y. Benjamini, Y. Hochberg, Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. R. Stat. Soc. Series B Stat. Methodol.* **57**, 289–300 (1995).

46. J. Liu, D. L. Silver, Founder cells shape brain evolution. *Cell*. **184**, 1965–1967 (2021).

47. McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University (Baltimore, MD), Online Mendelian Inheritance in Man, OMIM®. https://omim.org/.

48. J. den Hoed, E. de Boer, N. Voisin, A. J. M. Dingemans, N. Guex, L. Wiel, C. Nellaker, S. M.

562 Amudhavalli, S. Banka, F. S. Bena, B. Ben-Zeev, V. R. Bonagura, A.-L. Bruel, T. Brunet, H. G.
563 Brunner, H. B. Chew, J. Chrast, L. Cimbalistienė, H. Coon, DDD Study, E. C. Délot, F. Démurger,
564 A.-S. Denommé-Pichon, C. Depienne, D. Donnai, D. A. Dyment, O. Elpeleg, L. Faivre, C. Gilissen,
565 L. Granger, B. Haber, Y. Hachiya, Y. H. Abedi, J. Hanebeck, J. Y. Hehir-Kwa, B. Horist, T. Itai, A.
566 Jackson, R. Jewell, K. L. Jones, S. Joss, H. Kashii, M. Kato, A. A. Kattentidt-Mouravieva, F. Kok,
567 U. Kotzaeridou, V. Krishnamurthy, V. Kučinskas, A. Kuechler, A. Lavillaureix, P. Liu, L.
568 Manwaring, N. Matsumoto, B. Mazel, K. McWalter, V. Meiner, M. A. Mikati, S. Miyatake, T.
569 Mizuguchi, L. H. Moey, S. Mohammed, H. Mor-Shaked, H. Mountford, R. Newbury-Ecob, S.
570 Odent, L. Orec, M. Osmond, T. B. Palculict, M. Parker, A. K. Petersen, R. Pfundt, E. Preikšaitienė,
571 K. Radtke, E. Ranza, J. A. Rosenfeld, T. Santiago-Sim, C. Schwager, M. Sinnema, L. Snijders Blok,
572 R. C. Spillmann, A. P. A. Stegmann, I. Thiffault, L. Tran, A. Vaknin-Dembinsky, J. H. Vedovato-
573 Dos-Santos, S. A. Schrier Vergano, E. Vilain, A. Vitobello, M. Wagner, A. Waheeb, M. Willing, B.
574 Zuccarelli, U. Kini, D. F. Newbury, T. Kleefstra, A. Reymond, S. E. Fisher, L. E. L. M. Vissers,
575 Mutation-specific pathophysiological mechanisms define different neurodevelopmental disorders
576 associated with SATB1 dysfunction. *Am. J. Hum. Genet.* **108**, 346–356 (2021).

577 49. C. K. Bauer, J. R. Schwarz, Ether-à-go-go K+ channels: effective modulators of neuronal
578 excitability. *J. Physiol.* **596**, 769–783 (2018).

579 50. M. Borg Distefano, L. Hofstad Haugen, Y. Wang, H. Perdreau-Dahl, I. Kjos, D. Jia, J. P. Morth, J.
580 Neefjes, O. Bakke, C. Progida, TBC1D5 controls the GTPase cycle of Rab7b. *J. Cell Sci.* **131**
581 (2018).

582 51. S. H. Ridgway, R. H. Brownson, K. R. Van Alstyne, R. A. Hauser, Higher neuron densities in the
583 cerebral cortex and larger cerebellums may limit dive times of delphinids compared to deep-diving
584 toothed whales. *PLoS One*. **14**, e0226206 (2019).

585 52. D. Jeong, D. Lozano Casasbuenas, A. Gengatharan, K. Edwards, A. Saghatelyan, D. R. Kaplan, F.

D. Miller, S. A. Yuzwa, LRIG1-Mediated Inhibition of EGF Receptor Signaling Regulates Neural Precursor Cell Proliferation in the Neocortex. *Cell Rep.* **33**, 108257 (2020).

53. M. Á. Marqués-Torrejón, C. A. C. Williams, B. Southgate, N. Alfazema, M. P. Clements, C. Garcia-Diaz, C. Blin, N. Arranz-Emparan, J. Fraser, N. Gammoh, S. Parrinello, S. M. Pollard, LRIG1 is a gatekeeper to exit from quiescence in adult neural stem cells. *Nat. Commun.* **12**, 2594 (2021).

54. C. M. Deber, S. J. Reynolds, Central nervous system myelin: structure, function, and pathology. *Clin. Biochem.* **24**, 113–134 (1991).

55. P. Giusti-Rodríguez, L. Lu, Y. Yang, C. A. Crowley, X. Liu, I. Juric, J. S. Martin, A. Abnousi, S. Colby Allred, N. Ancalade, N. J. Bray, G. Breen, J. Bryois, C. M. Bulik, J. J. Crowley, J. Guintivano, P. R. Jansen, G. J. Jurjus, Y. Li, G. Mahajan, S. Marzi, J. Mill, M. C. O'Donovan, J. C. Overholser, M. J. Owen, A. F. Pardiñas, S. Pochareddy, D. Posthuma, G. Rajkowska, G. Santpere, J. E. Savage, N. Sestan, Y. Shin, C. A. Stockmeier, J. T. R. Walters, S. Yao, Bipolar Disorder Working Group of the Psychiatric Genomics Consortium, Eating Disorders Working Group of the Psychiatric Genomics Consortium, G. E. Crawford, F. Jin, M. Hu, Y. Li, P. F. Sullivan, https://www.biorxiv.org/content/10.1101/406330v2.

56. J. F. de Celis, R. Barrio, Regulation and function of Spalt proteins during animal development. *Int. J. Dev. Biol.* **53**, 1385–1398 (2009).

57. S. J. Harrison, M. Parrish, A. P. Monaghan, Sall3 is required for the terminal maturation of olfactory glomerular interneurons. *J. Comp. Neurol.* **507**, 1780–1794 (2008).

58. M. Parrish, T. Ott, C. Lance-Jones, G. Schuetz, A. Schwaeger-Nickolenko, A. P. Monaghan, Loss of the Sall3 gene leads to palate deficiency, abnormalities in cranial nerves, and perinatal lethality. *Mol. Cell. Biol.* **24**, 7102–7112 (2004).

59. M. Uhlén, L. Fagerberg, B. M. Hallström, C. Lindskog, P. Oksvold, A. Mardinoglu, Å. Sivertsson,

C. Kampf, E. Sjöstedt, A. Asplund, I. Olsson, K. Edlund, E. Lundberg, S. Navani, C. A.-K. Szigyarto, J. Odeberg, D. Djureinovic, J. O. Takanen, S. Hober, T. Alm, P.-H. Edqvist, H. Berling, H. Tegel, J. Mulder, J. Rockberg, P. Nilsson, J. M. Schwenk, M. Hamsten, K. von Feilitzen, M. Forsberg, L. Persson, F. Johansson, M. Zwahlen, G. von Heijne, J. Nielsen, F. Pontén, Proteomics. Tissue-based map of the human proteome. *Science*. **347**, 1260419 (2015).

60. A. J. Lawler, E. Ramamurthy, A. R. Brown, N. Shin, Y. Kim, N. Toong, I. M. Kaplow, M. Wirthlin, X. Zhang, G. Fox, A. R. Pfenning, https://www.biorxiv.org/content/10.1101/2021.04.15.439984v1.

61. A. Mo, E. A. Mukamel, F. P. Davis, C. Luo, G. L. Henry, S. Picard, M. A. Urich, J. R. Nery, T. J. Sejnowski, R. Lister, S. R. Eddy, J. R. Ecker, J. Nathans, Epigenomic Signatures of Neuronal Diversity in the Mammalian Brain. *Neuron*. **86**, 1369–1384 (2015).

62. B. Rudy, G. Fishell, S. Lee, J. Hjerling-Leffler, Three groups of interneurons account for nearly 100% of neocortical GABAergic neurons. *Dev. Neurobiol.* **71**, 45–61 (2011).

63. P. McColgan, J. Joubert, S. J. Tabrizi, G. Rees, The human motor cortex microcircuit: insights for neurodegenerative disease. *Nat. Rev. Neurosci.* **21**, 401–415 (2020).

64. G. Gonzalez-Burgos, R. Y. Cho, D. A. Lewis, Alterations in cortical network oscillations and parvalbumin neurons in schizophrenia. *Biol. Psychiatry*. **77**, 1031–1040 (2015).

65. S. Leimkühler, A. Freuer, J. A. S. Araujo, K. V. Rajagopalan, R. R. Mendel, Mechanistic studies of human molybdopterin synthase reaction and characterization of mutants identified in group B patients of molybdenum cofactor deficiency. *J. Biol. Chem.* **278**, 26127–26134 (2003).

66. E. Bayram, Y. Topcu, P. Karakaya, U. Yis, H. Cakmakci, K. Ichida, S. H. Kurul, Molybdenum cofactor deficiency: Review of 12 cases (MoCD and review). *Eur. J. Paediatr. Neurol.* **17**, 1–6 (2013).

631 67. T. Kröcher, I. Röckle, U. Diederichs, B. Weinhold, H. Burkhardt, Y. Yanagawa, R. Gerardy-Schahn,
632 H. Hildebrandt, A crucial role for polysialic acid in developmental interneuron migration and the
633 establishment of interneuron densities in the mouse prefrontal cortex. *Development*. **141**, 3022–3032
634 (2014).

635 68. Y. Curto, J. Alcaide, I. Röckle, H. Hildebrandt, J. Nacher, Effects of the Genetic Depletion of
636 Polysialyltransferases on the Structure and Connectivity of Interneurons in the Adult Prefrontal
637 Cortex. *Front. Neuroanat.* **13**, 6 (2019).

638 69. B. J. Crespi, T. L. Procyshyn, Williams syndrome deletions and duplications: Genetic windows to
639 understanding anxiety, sociality, autism, and schizophrenia. *Neurosci. Biobehav. Rev.* **79**, 14–26
640 (2017).

641 70. B. R. Ferguson, W.-J. Gao, PV Interneurons: Critical Regulators of E/I Balance for Prefrontal
642 Cortex-Dependent Behavior and Psychiatric Disorders. *Front. Neural Circuits*. **12**, 37 (2018).

643 71. M. Schwede, S. Nagpal, M. J. Gandal, N. N. Parikshak, K. Mirnics, D. H. Geschwind, E. M.
644 Morrow, Strong correlation of downregulated genes related to synaptic transmission and
645 mitochondria in post-mortem autism cerebral cortex. *J. Neurodev. Disord.* **10**, 18 (2018).

646 72. B. N. Phan, J. F. Bohlen, B. A. Davis, Z. Ye, H.-Y. Chen, B. Mayfield, S. R. Sripathy, S. Cerceo
647 Page, M. N. Campbell, H. L. Smith, D. Gallop, H. Kim, C. L. Thaxton, J. M. Simon, E. E. Burke, J.
648 H. Shin, A. J. Kennedy, J. D. Sweatt, B. D. Philpot, A. E. Jaffe, B. J. Maher, A myelin-related
649 transcriptomic profile is shared by Pitt–Hopkins syndrome models and human autism spectrum
650 disorder. *Nat. Neurosci.* **23**, 375–385 (2020).

651 73. B. C. Reiner, R. C. Crist, L. M. Stein, A. E. Weller, G. A. Doyle, G. Arauco-Shapiro, G. Turecki, T.
652 N. Ferraro, M. R. Hayes, W. H. Berrettini,
653 https://www.biorxiv.org/content/10.1101/2020.07.29.227355v3.

654 74. W. B. Ruzicka, S. Mohammadi, J. Davila-Velderrain, S. Subburaju, D. R. Tso, M. Hourihan, M.
655 Kellis, https://www.medrxiv.org/content/10.1101/2020.11.06.20225342v1.

656 75. A. L. Smith, E.-M. Jung, B. T. Jeon, W.-Y. Kim, Arid1b haploinsufficiency in parvalbumin- or
657 somatostatin-expressing interneurons leads to distinct ASD-like and ID-like behavior. *Sci. Rep.* **10**,
658 7834 (2020).

659 76. D. Lukas, T. H. Clutton-Brock, The evolution of social monogamy in mammals. *Science*. **341**, 526–
660 530 (2013).

661 77. Schizophrenia Working Group of the Psychiatric Genomics Consortium, S. Ripke, J. T. R. Walters,
662 M. C. O'Donovan, https://www.medrxiv.org/content/10.1101/2020.09.12.20192922v1.

663 78. E. Shuldiner, I. J. Koch, R. Y. Kartzinel, A. Hogan, L. Brubaker, S. Wanser, D. Stahler, C. D. L.
664 Wynne, E. A. Ostrander, J. S. Sinsheimer, Others, Structural variants in genes associated with
665 human Williams-Beuren syndrome underlie stereotypical hypersociability in domestic dogs. *Sci.*
666 *Adv.* **3**, e1700398 (2017).

667 79. M. Wirthlin, E. F. Chang, M. Knörnschild, L. A. Krubitzer, C. V. Mello, C. T. Miller, A. R.
668 Pfenning, S. C. Vernes, O. Tchernichovski, M. M. Yartsev, A Modular Approach to Vocal Learning:
669 Disentangling the Diversity of a Complex Behavioral Trait. *Neuron*. **104**, 87–99 (2019).

670 80. E. D. Jarvis, Learned birdsong and the neurobiology of human language. *Ann. N. Y. Acad. Sci.* **1016**,
671 749–777 (2004).

672 81. M. E. Wirthlin, X. Zhang, S. Annaldasula, I. M. Kaplow, D. E. Schaffer, A. J. Lawler, A. R. Brown,
673 M. Yartzev, A. P. Pfenning, Vocal learning-associated convergent evolution across mammals in
674 regulatory element and protein sequence. *Submitted to Science with Zoonomia Consoritum*.

675 82. D. Chabbert, X. Caubit, P. L. Roubertoux, M. Carlier, B. Habermann, B. Jacq, P. Salin, M. Metwaly,

C. Frahm, A. Fatmi, A. N. Garratt, D. Severac, E. Dubois, L. Kerkerian-Le Goff, L. Fasano, P. Gubellini, Postnatal Tshz3 Deletion Drives Altered Corticostriatal Function and Autism Spectrum Disorder–like Behavior. *Biol. Psychiatry*. **86**, 274–285 (2019).

83. S. S. P. Rao, M. H. Huntley, N. C. Durand, E. K. Stamenova, I. D. Bochkov, J. T. Robinson, A. L. Sanborn, I. Machol, A. D. Omer, E. S. Lander, E. L. Aiden, A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping. *Cell*. **159**, 1665–1680 (2014).

84. S. Volland, J. Esteve-Rudd, J. Hoo, C. Yee, D. S. Williams, A comparison of some organizational characteristics of the mouse central retina and the human macula. *PLoS One*. **10**, e0125631 (2015).

85. Y. Wu, H. Wang, H. Wang, E. A. Hadly, Rethinking the Origin of Primates by Reconstructing Their Diel Activity Patterns Using Genetics and Morphology. *Sci. Rep.* **7**, 11837 (2017).

86. M. T. Weirauch, A. Yang, M. Albu, A. G. Cote, A. Montenegro-Montero, P. Drewe, H. S. Najafabadi, S. A. Lambert, I. Mann, K. Cook, H. Zheng, A. Goity, H. van Bakel, J.-C. Lozano, M. Galli, M. G. Lewsey, E. Huang, T. Mukherjee, X. Chen, J. S. Reece-Hoyes, S. Govindarajan, G. Shaulsky, A. J. M. Walhout, F.-Y. Bouget, G. Ratsch, L. F. Larrondo, J. R. Ecker, T. R. Hughes, Determination and inference of eukaryotic transcription factor sequence specificity. *Cell*. **158**, 1431–1443 (2014).

87. T. L. Bailey, M. Boden, F. A. Buske, M. Frith, C. E. Grant, L. Clementi, J. Ren, W. W. Li, W. S. Noble, MEME Suite: Tools for motif discovery and searching. *Nucleic Acids Res.* **37**, W202–W208 (2009).

88. 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.* JN-RM-2534-20 (2021).

699 89. C. Liu, M. Wang, X. Wei, L. Wu, J. Xu, X. Dai, J. Xia, M. Cheng, Y. Yuan, P. Zhang, J. Li, T. Feng,
700 A. Chen, W. Zhang, F. Chen, Z. Shang, X. Zhang, B. A. Peters, L. Liu, An ATAC-seq atlas of
701 chromatin accessibility in mouse tissues. *Sci Data*. **6**, 65 (2019).

702 90. M. Wirthlin, I. M. Kaplow, A. J. Lawler, J. He, B. N. Phan, A. R. Brown, W. R. Stauffer, A. R.
703 Pfenning, https://www.biorxiv.org/content/10.1101/2020.10.27.356733v1.

704 91. A. R. Quinlan, I. M. Hall, BEDTools: a flexible suite of utilities for comparing genomic features.
705 *Bioinformatics*. **26**, 841–842 (2010).

706 92. M. M. Halstead, C. Kern, P. Saelao, Y. Wang, G. Chanthavixay, J. F. Medrano, A. L. Van
707 Eenennaam, I. Korf, C. K. Tuggle, C. W. Ernst, H. Zhou, P. J. Ross, A comparative analysis of
708 chromatin accessibility in cattle, pig, and mouse tissues. *BMC Genom.* **21**, 698 (2020).

709 93. J. B. Miesfeld, N. M. Ghiasvand, B. Marsh-Armstrong, N. Marsh-Armstrong, E. B. Miller, P. Zhang,
710 S. K. Manna, R. J. Zawadzki, N. L. Brown, T. Glaser, The Atoh7 remote enhancer provides
711 transcriptional robustness during retinal ganglion cell development. *Proc. Natl. Acad. Sci. U. S. A.*
712 **117**, 21690–21700 (2020).

713 94. T. J. Cherry, M. G. Yang, D. A. Harmin, P. Tao, A. E. Timms, M. Bauwens, R. Allikmets, E. M.
714 Jones, R. Chen, E. De Baere, M. E. Greenberg, Mapping the cis-regulatory architecture of the human
715 retina reveals noncoding genetic variation in disease. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 9001–9012
716 (2020).

717 95. Y. E. Li, S. Preissl, X. Hou, Z. Zhang, K. Zhang, R. Fang, Y. Qiu, O. Poirion, B. Li, H. Liu, X.
718 Wang, J. Y. Han, J. Lucero, Y. Yan, S. Kuan, D. Gorkin, M. Nunn, E. A. Mukamel, M. Margarita
719 Behrens, J. Ecker, B. Ren, An Atlas of Gene Regulatory Elements in Adult Mouse Cerebrum.
720 *Nature*. **598**, 129-136.

721 96. 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. Eggermont, 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).

97. R. H. Waterston, K. Lindblad-Toh, E. Birney, J. Rogers, J. F. Abril, P. Agarwal, R. Agarwala, R. Ainscough, M. Alexandersson, P. An, S. E. Antonarakis, J. Attwood, R. Baertsch, J. Bailey, K. Barlow, S. Beck, E. Berry, B. Birren, T. Bloom, P. Bork, M. Botcherby, N. Bray, M. R. Brent, D. G. Brown, S. D. Brown, C. Bult, J. Burton, J. Butler, R. D. Campbell, P. Carninci, S. Cawley, F. Chiaromonte, A. T. Chinwalla, D. M. Church, M. Clamp, C. Clee, F. S. Collins, L. L. Cook, R. R. Copley, A. Coulson, O. Couronne, J. Cuff, V. Curwen, T. Cutts, M. Daly, R. David, J. Davies, K. D. Delehaunty, J. Deri, E. T. Dermitzakis, C. Dewey, N. J. Dickens, M. Diekhans, S. Dodge, I. Dubchak, D. M. Dunn, S. R. Eddy, L. Elnitski, R. D. Emes, P. Eswara, E. Eyras, A. Felsenfeld, G. A. Fewell, P. Flicek, K. Foley, W. N. Frankel, L. A. Fulton, R. S. Fulton, T. S. Furey, D. Gage, R. A. Gibbs, G. Glusman, S. Gnerre, N. Goldman, L. Goodstadt, D. Grafham, T. A. Graves, E. D. Green, S. Gregory, R. Guigo, M. Guyer, R. C. Hardison, D. Haussler, Y. Hayashizaki, L. W. Hillier, A.

Hinrichs, W. Hlavina, T. Holzer, F. Hsu, A. Hua, T. Hubbard, A. Hunt, I. Jackson, D. B. Jaffe, L. S. Johnson, M. Jones, T. A. Jones, A. Joy, M. Kamal, E. K. Karlsson, D. Karolchik, A. Kasprzyk, J. Kawai, E. Keibler, C. Kells, W. J. Kent, A. Kirby, D. L. Kolbe, I. Korf, R. S. Kucherlapati, E. J. Kulbokas, D. Kulp, T. Landers, J. P. Leger, S. Leonard, I. Letunic, R. Levine, J. Li, M. Li, C. Lloyd, S. Lucas, B. Ma, D. R. Maglott, E. R. Mardis, L. Matthews, E. Mauceli, J. H. Mayer, M. McCarthy, W. R. McCombie, S. McLaren, K. McLay, J. D. McPherson, J. Meldrim, B. Meredith, J. P. Mesirov, W. Miller, T. L. Miner, E. Mongin, K. T. Montgomery, M. Morgan, R. Mott, J. C. Mullikin, D. M. Muzny, W. E. Nash, J. O. Nelson, M. N. Nhan, R. Nicol, Z. Ning, C. Nusbaum, M. J. O'Connor, Y. Okazaki, K. Oliver, E. O. Larty, L. Pachter, G. Parra, K. H. Pepin, J. Peterson, P. Pevzner, R. Plumb, C. S. Pohl, A. Poliakov, T. C. Ponce, C. P. Ponting, S. Potter, M. Quail, A. Reymond, B. A. Roe, K. M. Roskin, E. M. Rubin, A. G. Rust, R. Santos, V. Sapojnikov, B. Schultz, J. Schultz, M. S. Schwartz, S. Schwartz, C. Scott, S. Seaman, S. Searle, T. Sharpe, A. Sheridan, R. Shownkeen, S. Sims, J. B. Singer, G. Slater, A. Smit, D. R. Smith, B. Spencer, A. Stabenau, N. S. Strange-Thomann, C. Sugnet, M. Suyama, G. Tesler, J. Thompson, D. Torrents, E. Trevaskis, J. Tromp, C. Ucla, A. U. Vidal, J. P. Vinson, A. C. von Niederhausern, C. M. Wade, M. Wall, R. J. Weber, R. B. Weiss, M. C. Wendl, A. P. West, K. Wetterstrand, R. Wheeler, S. Whelan, J. Wierzbowski, D. Willey, S. Williams, R. K. Wilson, E. Winter, K. C. Worley, D. Wyman, S. Yang, S. P. Yang, E. M. Zdobnov, M. C. Zody, E. S. Lander, C. Mouse Genome Sequencing, Initial sequencing and comparative analysis of the mouse genome. *Nature*. **420**, 520–562 (2002).

98. B. Langmead, S. L. Salzberg, Fast gapped-read alignment with Bowtie 2. *Nat. Methods*. **9**, 357–359 (2012).

99. K. R. Rosenbloom, J. Armstrong, G. P. Barber, J. Casper, H. Clawson, M. Diekhans, T. R. Dreszer, P. A. Fujita, L. Guruvadoo, M. Haeussler, R. A. Harte, S. Heitner, G. Hickey, A. S. Hinrichs, R. Hubley, D. Karolchik, K. Learned, B. T. Lee, C. H. Li, K. H. Miga, N. Nguyen, B. Paten, B. J. Raney, A. F. A. Smit, M. L. Speir, A. S. Zweig, D. Haussler, R. M. Kuhn, W. J. Kent, The UCSC

Genome Browser database: 2015 update. *Nucleic Acids Res.* **43**, D670–D681 (2015).

100. H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. **25**, 2078–2079 (2009).

101. F. Ramírez, D. P. Ryan, B. Grüning, V. Bhardwaj, F. Kilpert, A. S. Richter, S. Heyne, F. Dündar, T. Manke, deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* **44**, W160–5 (2016).

102. 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).

103. BRAIN Initiative Cell Census Network (BICCN), A multimodal cell census and atlas of the mammalian primary motor cortex. *Nature*. **598**, 86-102 (2021).

104. G. K. Marinov, A. Kundaje, P. Park B. Wold, Large-scale quality analysis of published ChIP-seq data. *G3 (Bethesda)*. **4**, 209-223 (2013).

105. Q. Li, J. B. Brown, H. Huang, P. J. Bickel, Measuring reproducibility of high-throughput experiments. *Ann. Appl. Stat.* **5**, 1752–1779 (2011).

106. 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. Uszczynska-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.

Reymond, M. L. Tress, P. Flicek, GENCODE reference annotation for the human and mouse genomes. *Nucleic Acids Res.* **47**, D766–D773 (2019).

107. N. A. O'Leary, M. W. Wright, J. R. Brister, S. Ciufo, D. Haddad, R. McVeigh, B. Rajput, B. Robbertse, B. Smith-White, D. Ako-Adjei, A. Astashyn, A. Badretdin, Y. Bao, O. Blinkova, V. Brover, V. Chetvernin, J. Choi, E. Cox, O. Ermolaeva, C. M. Farrell, T. Goldfarb, T. Gupta, D. Haft, E. Hatcher, W. Hlavina, V. S. Joardar, V. K. Kodali, W. Li, D. Maglott, P. Masterson, K. M. McGarvey, M. R. Murphy, K. O'Neill, S. Pujar, S. H. Rangwala, D. Rausch, L. D. Riddick, C. Schoch, A. Shkeda, S. S. Storz, H. Sun, F. Thibaud-Nissen, I. Tolstoy, R. E. Tully, A. R. Vatsan, C. Wallin, D. Webb, W. Wu, M. J. Landrum, A. Kimchi, T. Tatusova, M. DiCuccio, P. Kitts, T. D. Murphy, K. D. Pruitt, Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.* **44**, D733–D745 (2016).

108. W. J. Kent, R. Baertsch, A. Hinrichs, W. Miller, D. Haussler, Evolution's cauldron: duplication, deletion, and rearrangement in the mouse and human genomes. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 11484–11489 (2003).

109. D. Jebb, Z. Huang, M. Pippel, G. M. Hughes, K. Lavrichenko, P. Devanna, S. Winkler, L. S. Jermiin, 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. Puechmaille, 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).

110. Y. Zhang, T. Liu, C. A. Meyer, J. Eeckhoute, D. S. Johnson, B. E. Bernstein, C. Nussbaum, R. M. Myers, M. Brown, W. Li, X. S. Liu, Model-based Analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).

111. M. Ghandi, M. Mohammad-Noori, N. Ghareghani, D. Lee, L. Garraway, M. A. Beer, gkmSVM: an R package for gapped-kmer SVM. *Bioinformatics*. **32**, 2205–2207 (2016).

112. W. Huber, V. J. Carey, R. Gentleman, S. Anders, M. Carlson, B. S. Carvalho, H. C. Bravo, S. Davis, L. Gatto, T. Girke, R. Gottardo, F. Hahne, K. D. Hansen, R. A. Irizarry, M. Lawrence, M. I. Love, J. MacDonald, V. Obenchain, A. K. Oleś, H. Pagès, A. Reyes, P. Shannon, G. K. Smyth, D. Tenenbaum, L. Waldron, M. Morgan, Orchestrating high-throughput genomic analysis with Bioconductor. *Nat. Methods*. **12**, 115–121 (2015).

113. A. Khan, R. Riudavets Puig, P. Boddie, A. Mathelier, BiasAway: command-line and web server to generate nucleotide composition-matched DNA background sequences. *Bioinformatics*. **37**, 1607–1609 (2021).

114. R. Worsley Hunt, A. Mathelier, L. Del Peso, W. W. Wasserman, Improving analysis of transcription factor binding sites within ChIP-Seq data based on topological motif enrichment. *BMC Genom*. **15**, 472 (2014).

115. B. Alipanahi, A. Delong, M. T. Weirauch, B. J. Frey, Predicting the sequence specificities of DNA- and RNA-binding proteins by deep learning. *Nat. Biotechnol.* **33**, 831–838 (2015).

116. F. Chollet, Keras (2015). https://keras.io.

117. J. Bergstra, O. Breuleux, F. Bastien, P. Lamblin, R. Pascanu, G. Desjardins, J. Turian, D. Warde-Farley, Y. Bengio, Theano: A CPU and GPU Math Compiler in Python. *Proceedings of the Python for scientific computing conference (SciPy)*. **4**, 1–7 (2010).

118. M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, R. Jozefowicz, Y. Jia, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, M. Schuster, R. Monga, S. Moore, D. Murray, C. Olah, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F.

Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, X. Zheng, TensorFlow: a system for large-scale machine learning. *Proceedings of the 12th USENIX conference on Operating Systems Design and Implementation*. **12**, 262-283 (2016).

119. F. Pedregosa, G. Varoquaux, Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).

120. J. Grau, I. Grosse, J. Keilwagen, PRROC: Computing and visualizing Precision-recall and receiver operating characteristic curves in R. *Bioinformatics*. **31**, 2595–2597 (2015).

121. K. He, X. Zhang, S. Ren, J. Sun, Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification. *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition*. 1026–1034 (2015).

122. X. Glorot, Y. Bengio, Understanding the difficulty of training deep feedforward neural networks. *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics, Proceedings of Machine Learning Research*. **9**, 249–256 (2010).

123. L. N. Smith, Cyclical Learning Rates for Training Neural Networks. *2017 IEEE Winter Conference on Applications of Computer Vision (WACV)*. 464–472 (2017).

124. A. Shrikumar, K. Tian, A. Shcherbina, Ž. Avsec, A. Banerjee, M. Sharmin, S. Nair, A. Kundaje, https://arxiv.org/abs/1811.00416.

125. S. Gupta, J. A. Stamatoyannopoulos, T. L. Bailey, W. Noble, Quantifying similarity between motifs. *Genome Biol.* **8**, R24 (2007).

126. I. V. Kulakovskiy, I. E. Vorontsov, I. S. Yevshin, R. N. Sharipov, A. D. Fedorova, E. I. Rumynskiy, Y. A. Medvedeva, A. Magana-Mora, V. B. Bajic, D. A. Papatsenko, F. A. Kolpakov, V. J. Makeev, HOCOMOCO: towards a complete collection of transcription factor binding models for

863 human and mouse via large-scale ChIP-Seq analysis. *Nucleic Acids Res.* **46**, D252–D259 (2018).

864 127. A. Kowalczyk, R. Partha, N. L. Clark, M. Chikina, Pan-mammalian analysis of molecular
865 constraints underlying extended lifespan. *Elife*. **9**, e51089 (2020).

866 128. M. Chikina, J. D. Robinson, N. L. Clark, Hundreds of Genes Experienced Convergent Shifts in
867 Selective Pressure in Marine Mammals. *Mol. Biol. Evol.* **33**, 2182–2192 (2016).

868 129. B. Kirilenko, Zoonomia Consortium, M. Hiller, Machine-learning approach to infer orthologs and
869 integrate gene annotation with orthology inference at scale. *Submitted to Science with Zoonomia*
870 *Consoritum.*

871 130. R Core Team, R: A language and environment for statistical computing. R Foundation for
872 statistical computing (2021). https://www.R-project.org/.

873 131. P. Machanick, T. L. Bailey, MEME-ChIP: Motif analysis of large DNA datasets. *Bioinformatics*.
874 **27**, 1696–1697 (2011).

875 132. N. M. Foley, V. C. Mason, A. J. Harris, K. R. Bredemeyer, J. Damas, H. A. Lewin, E. Eizirik, J.
876 Gatesy, Zoonomia Consortium, M. S. Springer, W. J. Murphy, A genomic timescale for placental
877 mammal evolution. *Submitted to Science with Zoonomia Consoritum*.

878 133. D. Eddelbuettel, R. François, J. Allaire, K. Ushey, Q. Kou, N. Russel, J. Chambers, D. Bates,
879 Rcpp: Seamless R and C++ integration. *J. Stat. Softw.* **40**, 1–18 (2011).

880 134. D. Yekutieli, Y. Benjamini, Resampling-based false discovery rate controlling multiple test
881 procedures for correlated test statistics. *J. Stat. Plan. Inference*. **82**, 171–196 (1999).

882 135. E. Paradis, K. Schliep, ape 5.0: an environment for modern phylogenetics and evolutionary
883 analyses in R. *Bioinformatics*. **35**, 526–528 (2019).

884 136. M. W. Pennell, J. M. Eastman, G. J. Slater, J. W. Brown, J. C. Uyeda, R. G. FitzJohn, M. E.
885 Alfaro, L. J. Harmon, geiger v2.0: an expanded suite of methods for fitting macroevolutionary
886 models to phylogenetic trees. *Bioinformatics*. **30**, 2216-2218 (2014).

887 137. N. Ignatiadis, B. Klaus, J. B. Zaugg, W. Huber, Data-driven hypothesis weighting increases
888 detection power in genome-scale multiple testing. *Nat. Methods*. **13**, 577–580 (2016).

889 138. J. Harrow, A. Frankish, J. M. Gonzalez, E. Tapanari, M. Diekhans, F. Kokocinski, B. L. Aken, D.
890 Barrell, A. Zadissa, S. Searle, I. Barnes, A. Bignell, V. Boychenko, T. Hunt, M. Kay, G. Mukherjee,
891 J. Rajan, G. Despacio-Reyes, G. Saunders, C. Steward, R. Harte, M. Lin, C. Howald, A. Tanzer, T.
892 Derrien, J. Chrast, N. Walters, S. Balasubramanian, B. Pei, M. Tress, J. M. Rodriguez, I. Ezkurdia, J.
893 Van Baren, M. Brent, D. Haussler, M. Kellis, A. Valencia, A. Reymond, M. Gerstein, R. Guigó, T. J.
894 Hubbard, GENCODE: The reference human genome annotation for the ENCODE project. *Genome*
895 *Res.* **22**, 1760–1774 (2012).

896 139. K. Watanabe, E. Taskesen, A. van Bochoven, D. Posthuma, Functional mapping and annotation
897 of genetic associations with FUMA. *Nat. Commun.* **8**, 1826 (2017).

898 140. K. Yamada, D. J. Gerber, Y. Iwayama, T. Ohnishi, H. Ohba, T. Toyota, J. Aruga, Y. Minabe, S.
899 Tonegawa, T. Yoshikawa, Genetic analysis of the calcineurin pathway identifies members of the
900 EGR gene family, specifically EGR3, as potential susceptibility candidates in schizophrenia. *Proc.*
901 *Natl. Acad. Sci. U. S. A.* **104**, 2815–2820 (2007).

902 141. J. M. Olson, A. Asakura, L. Snider, R. Hawkes, A. Strand, J. Stoeck, A. Hallahan, J. Pritchard, S.
903 J. Tapscott, NeuroD2 is necessary for development and survival of central nervous system neurons.
904 *Dev. Biol.* **234**, 174–187 (2001).

905 142. T. Herdegen, J. D. Leah, Inducible and constitutive transcription factors in the mammalian
906 nervous system: control of gene expression by Jun, Fos and Krox, and CREB/ATF proteins. *Brain*

907 *Res. Brain Res. Rev.* **28**, 370–490 (1998).

908 143. K. S. Wendt, K. Yoshida, T. Itoh, M. Bando, B. Koch, E. Schirghuber, S. Tsutsumi, G. Nagae, K.
909 Ishihara, T. Mishiro, K. Yahata, F. Imamoto, H. Aburatani, M. Nakao, N. Imamoto, K. Maeshima,
910 K. Shirahige, J. M. Peters, Cohesin mediates transcriptional insulation by CCCTC-binding factor.
911 *Nature*. **451**, 796–801 (2008).

912 144. L. Isbel, L. Prokopuk, H. Wu, L. Daxinger, H. Oey, A. Spurling, A. J. Lawther, M. W. Hale, E.
913 Whitelaw, Wiz binds active promoters and CTCF-binding sites and is required for normal behaviour
914 in the mouse. *Elife*. **5**, e15082 (2016).

915 145. D. Zhang, D. C. Zeldin, P. J. Blackshear, Regulatory factor X4 variant 3: a transcription factor
916 involved in brain development and disease. *J. Neurosci. Res.* **85**, 3515–3522 (2007).

917 146. P. Xu, J. P. Morrison, J. F. Foley, D. J. Stumpo, T. Ward, D. C. Zeldin, P. J. Blackshear,
918 Conditional ablation of the RFX4 isoform 1 transcription factor: Allele dosage effects on brain
919 phenotype. *PLoS One*. **13**, e0190561 (2018).

920 147. H. Taniguchi, S. Okamuro, M. Koji, T. Waku, K. Kubo, A. Hatanaka, Y. Sun, A. M. M. A.
921 Chowdhury, A. Fukamizu, A. Kobayashi, Possible roles of the transcription factor Nrf1 (NFE2L1) in
922 neural homeostasis by regulating the gene expression of deubiquitinating enzymes. *Biochem.*
923 *Biophys. Res. Commun.* **484**, 176–183 (2017).

924 148. F. Gachon, P. Fonjallaz, F. Damiola, P. Gos, T. Kodama, J. Zakany, D. Duboule, B. Petit, M.
925 Tafti, U. Schibler, The loss of circadian PAR bZip transcription factors results in epilepsy. *Genes*
926 *Dev.* **18**, 1397–1412 (2004).

927 149. C. Chen, G. A. Lee, A. Pourmorady, E. Sock, M. J. Donoghue, Orchestration of Neuronal
928 Differentiation and Progenitor Pool Expansion in the Developing Cortex by SoxC Genes. *J.*
929 *Neurosci.* **35**, 10629–10642 (2015).

150. S. Shim, K. Y. Kwan, M. Li, V. Lefebvre, N. Sestan, Cis-regulatory control of corticospinal system development and evolution. *Nature*. **486**, 74–79 (2012).

151. T. Hisaoka, Y. Nakamura, E. Senba, Y. Morikawa, The forkhead transcription factors, Foxp1 and Foxp2, identify different subpopulations of projection neurons in the mouse cerebral cortex. *Neuroscience*. **166**, 551–563 (2010).

152. K. Ohi, R. Hashimoto, Y. Yasuda, M. Kiribayashi, N. Iike, T. Yoshida, M. Azechi, K. Ikezawa, H. Takahashi, T. Morihara, R. Ishii, S. Tagami, M. Iwase, M. Okochi, K. Kamino, H. Kazui, T. Tanaka, T. Kudo, M. Takeda, TATA box-binding protein gene is associated with risk for schizophrenia, age at onset and prefrontal function. *Genes Brain Behav.* **8**, 473–480 (2009).

153. B. Muralidharan, Z. Khatri, U. Maheshwari, R. Gupta, B. Roy, S. J. Pradhan, K. Karmodiya, H. Padmanabhan, A. S. Shetty, C. Balaji, U. Kolthur-Seetharam, J. D. Macklis, S. Galande, S. Tole, LHX2 Interacts with the NuRD Complex and Regulates Cortical Neuron Subtype Determinants Fezf2 and Sox11. *J. Neurosci.* **37**, 194–203 (2017).

154. H. Yang, J. Kim, Y. Kim, S.-W. Jang, N. Sestan, S. Shim, Cux2 expression regulated by Lhx2 in the upper layer neurons of the developing cortex. *Biochem. Biophys. Res. Commun.* **521**, 874–879 (2020).

155. The Human Protein Atlas. http://www.proteinatlas.org.

156. Y. Xue, C. Guo, F. Hu, W. Zhu, S. Mao, PPARA/RXRA signalling regulates the fate of hepatic non-esterified fatty acids in a sheep model of maternal undernutrition. *Biochim. Biophys. Acta Mol. Cell Biol. Lipids.* **1865**, 158548 (2020).

157. F. Stossi, R. D. Dandekar, H. Johnson, P. Lavere, C. E. Foulds, M. G. Mancini, M. A. Mancini, Tributyltin chloride (TBT) induces RXRA down-regulation and lipid accumulation in human liver cells. *PLoS One*. **14**, e0224405 (2019).

158. K. J. Wangensteen, S. Zhang, L. E. Greenbaum, K. H. Kaestner, A genetic screen reveals Foxa3 and TNFR1 as key regulators of liver repopulation. *Genes Dev.* **29**, 904–909 (2015).

159. J. R. Friedman, K. H. Kaestner, The Foxa family of transcription factors in development and metabolism. *Cell. Mol. Life Sci.* **63**, 2317–2328 (2006).

160. H. Schrem, J. Klempnauer, J. Borlak, Liver-enriched transcription factors in liver function and development. Part II: the C/EBPs and D site-binding protein in cell cycle control, carcinogenesis, circadian gene regulation, liver regeneration, apoptosis, and liver-specific gene regulation. *Pharmacol. Rev.* **56**, 291–330 (2004).

161. P. Hatzis, I. Kyrmizi, I. Talianidis, Mitogen-activated protein kinase-mediated disruption of enhancer-promoter communication inhibits hepatocyte nuclear factor 4alpha expression. *Mol. Cell. Biol.* **26**, 7017–7029 (2006).

162. J. B. Nielsen, O. Rom, I. Surakka, S. E. Graham, W. Zhou, T. Roychowdhury, L. G. Fritsche, S. A. Gagliano Taliun, C. Sidore, Y. Liu, M. E. Gabrielsen, A. H. Skogholt, B. Wolford, W. Overton, Y. Zhao, J. Chen, H. Zhang, W. E. Hornsby, A. Acheampong, A. Grooms, A. Schaefer, G. J. M. Zajac, L. Villacorta, J. Zhang, B. Brumpton, M. Løset, V. Rai, P. R. Lundegaard, M. S. Olesen, K. D. Taylor, N. D. Palmer, Y.-D. Chen, S. H. Choi, S. A. Lubitz, P. T. Ellinor, K. C. Barnes, M. Daya, N. Rafaels, S. T. Weiss, J. Lasky-Su, R. P. Tracy, R. S. Vasan, L. A. Cupples, R. A. Mathias, L. R. Yanek, L. C. Becker, P. A. Peyser, L. F. Bielak, J. A. Smith, S. Aslibekyan, B. A. Hidalgo, D. K. Arnett, M. R. Irvin, J. G. Wilson, S. K. Musani, A. Correa, S. S. Rich, X. Guo, J. I. Rotter, B. A. Konkle, J. M. Johnsen, A. E. Ashley-Koch, M. J. Telen, V. A. Sheehan, J. Blangero, J. E. Curran, J. M. Peralta, C. Montgomery, W. H.-H. Sheu, R.-H. Chung, K. Schwander, S. M. Nouraie, V. R. Gordeuk, Y. Zhang, C. Kooperberg, A. P. Reiner, R. D. Jackson, E. R. Bleecker, D. A. Meyers, X. Li, S. Das, K. Yu, J. LeFaive, A. Smith, T. Blackwell, D. Taliun, S. Zollner, L. Forer, S. Schoenherr, C. Fuchsberger, A. Pandit, M. Zawistowski, S. Kheterpal, C. M. Brummett, P. Natarajan, D.

977 Schlessinger, S. Lee, H. M. Kang, F. Cucca, O. L. Holmen, B. O. Åsvold, M. Boehnke, S.
978 Kathiresan, G. R. Abecasis, Y. E. Chen, C. J. Willer, K. Hveem, Loss-of-function genomic variants
979 highlight potential therapeutic targets for cardiovascular disease. *Nat. Commun.* **11**, 6417 (2020).

980 163. J. E. Phillipa, V. G. Corces, CTCF: Master Weaver of the Genome. *Cell*. **137**, 1194–1211 (2009).

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

985 165. T. Medici, P. J. Shortland, Effects of peripheral nerve injury on parvalbumin expression in adult
986 rat dorsal root ganglion neurons. *BMC Neurosci.* **16**, 93 (2015).

987 166. L. Lim, D. Mi, A. Llorca, O. Marín, Development and Functional Diversification of Cortical
988 Interneurons. *Neuron*. **100**, 294–313 (2018).

989 167. E. L.-L. Pai, J. Chen, S. Fazel Darbandi, F. S. Cho, J. Chen, S. Lindtner, J. S. Chu, J. T. Paz, D.
990 Vogt, M. F. Paredes, J. L. Rubenstein, Maf and Mafb control mouse pallial interneuron fate and
991 maturation through neuropsychiatric disease gene regulation. *Elife*. **9**, e54903 (2020).

992 168. A. M. Barbieri, G. Lupo, A. Bulfone, M. Andreazzoli, M. Mariani, F. Fougerousse, G. G.
993 Consalez, G. Borsani, J. S. Beckmann, G. Barsacchi, A. Ballabio, S. Banfi, A homeobox gene, vax2,
994 controls the patterning of the eye dorsoventral axis. *Proc. Natl. Acad. Sci. U. S. A.* **96**, 10729–10734
995 (1999).

996 169. G. Alfano, A. Z. Shah, G. Jeffery, S. S. Bhattacharya, First insights into the expression of VAX2
997 in humans and its localization in the adult primate retina. *Exp. Eye Res.* **148**, 24–29 (2016).

998 170. J. Li, J.-S. Zhang, J.-Y. Zhao, G.-G. Han, Role of Smad4 from ocular surface ectoderm in retinal

vasculature development. *Int. J. Ophthalmol.* **13**, 231–238 (2020).

171. R. Ratnapriya, O. A. Sosina, M. R. Starostik, M. Kwicklis, R. J. Kapphahn, L. G. Fritsche, A. Walton, M. Arvanitis, L. Gieser, A. Pietraszkiewicz, S. R. Montezuma, E. Y. Chew, A. Battle, G. R. Abecasis, D. A. Ferrington, N. Chatterjee, A. Swaroop, Retinal transcriptome and eQTL analyses identify genes associated with age-related macular degeneration. *Nat. Genet.* **51**, 606–610 (2019).

172. A.-G. Wang, C.-H. Chen, C.-W. Yang, M.-Y. Yen, W.-M. Hsu, J.-H. Liu, M.-J. Fann, Change of gene expression profiles in the retina following optic nerve injury. *Brain Res. Mol. Brain Res.* **101**, 82–92 (2002).

173. A. Swaroop, D. Kim, D. Forrest, Transcriptional regulation of photoreceptor development and homeostasis in the mammalian retina. *Nat. Rev. Neurosci.* **11**, 563–576 (2010).

174. P.-H. Fabre, L. Hautier, D. Dimitrov, E. J. P. Douzery, A glimpse on the pattern of rodent diversification: a phylogenetic approach. *BMC Evol. Biol.* **12**, 88 (2012).

175. T. C. Giarla, J. A. Esselstyn, The Challenges of Resolving a Rapid, Recent Radiation: Empirical and Simulated Phylogenomics of Philippine Shrews. *Syst. Biol*. **64**, 727-740 (2015).

176. G. Shi, Z. Zhang, Rap2B promotes the proliferation and migration of human glioma cells via activation of the ERK pathway. *Oncol. Lett.* **21**, 314 (2021).

177. S. P. F. Ensign, A. Roos, I. T. Mathews, H. D. Dhruv, S. Tuncali, J. N. Sarkaria, M. H. Symons, J. C. Loftus, M. E. Berens, N. L. Tran, SGEF Is Regulated via TWEAK/Fn14/NF-κB Signaling and Promotes Survival by Modulation of the DNA Repair Response to Temozolomide. *Mol. Cancer Res.* **14**, 302–312 (2016).

178. H. Liu, Y.-N. Lu, T. Paul, G. Periz, M. T. Banco, A. R. Ferré-D'Amaré, J. D. Rothstein, L. R. Hayes, S. Myong, J. Wang, A Helicase Unwinds Hexanucleotide Repeat RNA G-Quadruplexes and

Facilitates Repeat-Associated Non-AUG Translation. *J. Am. Chem. Soc.* **143**, 7368–7379 (2021).

179. Y. Shangguan, X. Xu, B. Ganbat, Y. Li, W. Wang, Y. Yang, X. Lu, C. Du, X. Tian, X. Wang, CNTNAP4 Impacts Epilepsy Through GABAA Receptors Regulation: Evidence From Temporal Lobe Epilepsy Patients and Mouse Models. *Cereb. Cortex*. **28**, 3491–3504 (2018).

180. F.-T. Yin, T. Futagawa, D. Li, Y.-X. Ma, M.-H. Lu, L. Lu, S. Li, Y. Chen, Y.-J. Cao, Z. Z. Yang, S. Oiso, K. Nishida, S. Kuchiiwa, K. Watanabe, K. Yamada, Y. Takeda, Z.-C. Xiao, Q.-H. Ma, Caspr4 interaction with LNX2 modulates the proliferation and neuronal differentiation of mouse neural progenitor cells. *Stem Cells Dev.* **24**, 640–652 (2015).

181. C. Sun, J. Song, Y. Jiang, C. Zhao, J. Lu, Y. Li, Y. Wang, M. Gao, J. Xi, S. Luo, M. Li, K. Donaldson, S. N. Oprescu, T. P. Slavin, S. Lee, P. L. Magoulas, A. M. Lewis, L. Emrick, S. R. Lalani, Z. Niu, M. L. Landsverk, M. Walkiewicz, R. E. Person, H. Mei, J. A. Rosenfeld, Y. Yang, A. Antonellis, Y.-M. Hou, J. Lin, V. W. Zhang, Loss-of-function mutations in Lysyl-tRNA synthetase cause various leukoencephalopathy phenotypes. *Neurol Genet*. **5**, e565 (2019).

182. H. J. McMillan, P. Humphreys, A. Smith, J. Schwartzentruber, P. Chakraborty, D. E. Bulman, C. L. Beaulieu, FORGE Canada Consortium, J. Majewski, K. M. Boycott, M. T. Geraghty, Congenital Visual Impairment and Progressive Microcephaly Due to Lysyl-Transfer Ribonucleic Acid (RNA) Synthetase (KARS) Mutations: The Expanding Phenotype of Aminoacyl-Transfer RNA Synthetase Mutations in Human Disease. *J. Child Neurol.* **30**, 1037–1043 (2015).

183. C. C. Hong, A. T. Tang, M. R. Detter, J. P. Choi, R. Wang, X. Yang, A. A. Guerrero, C. F. Wittig, N. Hobson, R. Girard, R. Lightle, T. Moore, R. Shenkar, S. P. Polster, L. M. Goddard, A. A. Ren, N. A. Leu, S. Sterling, J. Yang, L. Li, M. Chen, P. Mericko-Ishizuka, L. E. Dow, H. Watanabe, M. Schwaninger, W. Min, D. A. Marchuk, X. Zheng, I. A. Awad, M. L. Kahn, Cerebral cavernous malformations are driven by ADAMTS5 proteolysis of versican. *J. Exp. Med.* **217**, e20200140

(2020).

184. H. Kumon, Y. Yoshino, Y. Funahashi, H. Mori, M. Ueno, Y. Ozaki, K. Yamazaki, S. Ochi, T. Mori, J.-I. Iga, M. Nagai, M. Nomoto, S.-I. Ueno, PICALM mRNA Expression in the Blood of Patients with Neurodegenerative Diseases and Geriatric Depression. *J. Alzheimers. Dis.* **79**, 1055–1062 (2021).

185. E. Cooney, W. Bi, A. E. Schlesinger, S. Vinson, L. Potocki, Novel EED mutation in patient with Weaver syndrome. *Am. J. Med. Genet. A.* **173**, 541–545 (2017).

186. J. Wang, L. Yang, C. Dong, J. Wang, L. Xu, Y. Qiu, Q. Weng, C. Zhao, M. Xin, Q. R. Lu, EED-mediated histone methylation is critical for CNS myelination and remyelination by inhibiting WNT, BMP, and senescence pathways. *Sci Adv*. **6**, eaaz6477 (2020).

187. G. Helman, A. Zerem, A. Almad, J. L. Hacker, S. Woidill, S. Sase, A. N. LeFevre, J. Ekstein, M. M. Johansson, C. A. Stutterd, R. J. Taft, C. Simons, J. B. Grinspan, A. Pizzino, J. L. Schmidt, B. Harding, Y. Hirsch, A. N. Viaene, A. Fattal-Valevski, A. Vanderver, Further Delineation of the Clinical and Pathologic Features of HIKESHI-Related Hypomyelinating Leukodystrophy. *Pediatr. Neurol.* **121**, 11–19 (2021).

188. S. Edvardson, S. Kose, C. Jalas, A. Fattal-Valevski, A. Watanabe, Y. Ogawa, H. Mamada, A. M. Fedick, S. Ben-Shachar, N. R. Treff, A. Shaag, S. Bale, J. Gärtner, N. Imamoto, O. Elpeleg, Leukoencephalopathy and early death associated with an Ashkenazi-Jewish founder mutation in the Hikeshi gene. *J. Med. Genet.* **53**, 132–137 (2016).

189. A. Oguro-Ando, R. A. Bamford, W. Sital, J. J. Sprengers, A. Zuko, J. M. Matser, H. Oppelaar, A. Sarabdjitsingh, M. Joëls, J. P. H. Burbach, M. J. Kas, Cntn4, a risk gene for neuropsychiatric disorders, modulates hippocampal synaptic plasticity and behavior. *Transl. Psychiatry*. **11**, 106 (2021).

190. J. Hu, J. Liao, M. Sathanoori, S. Kochmar, J. Sebastian, S. A. Yatsenko, U. Surti, CNTN6 copy number variations in 14 patients: a possible candidate gene for neurodevelopmental and neuropsychiatric disorders. *J. Neurodev. Disord.* **7**, 26 (2015).

191. L. M. Hansford, S. A. Smith, M. Haber, M. D. Norris, B. Cheung, G. M. Marshall, Cloning and characterization of the human neural cell adhesion molecule, CNTN4 (alias BIG-2). *Cytogenet. Genome Res.* **101**, 17–23 (2003).

192. Y. Kamei, O. Tsutsumi, Y. Taketani, K. Watanabe, cDNA cloning and chromosomal localization of neural adhesion molecule NB-3 in human. *J. Neurosci. Res.* **51**, 275–283 (1998).

193. K. Baker, S. L. Gordon, H. Melland, F. Bumbak, D. J. Scott, T. J. Jiang, D. Owen, B. J. Turner, S. G. Boyd, M. Rossi, M. Al-Raqad, O. Elpeleg, D. Peck, G. M. S. Mancini, M. Wilke, M. Zollino, G. Marangi, H. Weigand, I. Borggraefe, T. Haack, Z. Stark, S. Sadedin, Broad Center for Mendelian Genomics, T. Y. Tan, Y. Jiang, R. A. Gibbs, S. Ellingwood, M. Amaral, W. Kelley, M. A. Kurian, M. A. Cousin, F. L. Raymond, SYT1-associated neurodevelopmental disorder: a case series. *Brain*. **141**, 2576–2591 (2018).

194. A. Shukla, R. P. Saneto, M. Hebbar, G. Mirzaa, K. M. Girisha, A neurodegenerative mitochondrial disease phenotype due to biallelic loss-of-function variants in PNPLA8 encoding calcium-independent phospholipase A2γ. *Am. J. Med. Genet. A.* **176**, 1232–1237 (2018).

195. M. Paucar, R. Ågren, T. Li, S. Lissmats, Å. Bergendal, J. Weinberg, D. Nilsson, I. Savichetva, K. Sahlholm, J. Nilsson, P. Svenningsson, V374A KCND3 Pathogenic Variant Associated With Paroxysmal Ataxia Exacerbations. *Neurol Genet*. **7**, e546 (2021).

196. G. Skariah, K. J. Perry, J. Drnevich, J. J. Henry, S. Ceman, RNA helicase Mov10 is essential for gastrulation and central nervous system development. *Dev. Dyn.* **247**, 660–671 (2018).

197. Q. Xiao, M. Dong, F. Cheng, F. Mao, W. Zong, K. Wu, H. Wang, R. Xie, B. Wang, T. Lei, D.

Guo, LRIG2 promotes the proliferation and cell cycle progression of glioblastoma cells in vitro and in vivo through enhancing PDGFRβ signaling. *Int. J. Oncol.* **53**, 1069–1082 (2018).

198. A.-R. Lee, K. W. Ko, H. Lee, Y.-S. Yoon, M.-R. Song, C.-S. Park, Putative Cell Adhesion Membrane Protein Vstm5 Regulates Neuronal Morphology and Migration in the Central Nervous System. *J. Neurosci.* **36**, 10181–10197 (2016).

199. A. Fattal-Valevski, L. Ben Sira, T. Lerman-Sagie, R. Strausberg, A. Bloch-Mimouni, S. Edvardson, R. Kaufman, V. Chernuha, N. Schneebaum Sender, G. Heimer, B. Ben Zeev, Delineation of the phenotype of MED17-related disease in Caucasus-Jewish families. *Eur. J. Paediatr. Neurol.* **32**, 40–45 (2021).

200. I. R. Raslan, P. C. A. de Assis Pereira Matos, V. Boaratti Ciarlariello, K. H. Daghastanli, A. B. R. Rosa, J. H. Arita, C. S. Aranda, O. G. P. Barsottini, J. L. Pedroso, Beyond Typical Ataxia Telangiectasia: How to Identify the Ataxia Telangiectasia-Like Disorders. *Mov Disord Clin Pract*. **8**, 118–125 (2021).

201. C. S. Subhramanyam, Q. Cao, C. Wang, Z. S. L. Heng, Z. Zhou, Q. Hu, Role of PIWI-like 4 in modulating neuronal differentiation from human embryonal carcinoma cells. *RNA Biol.* **17**, 1613–1624 (2020).

202. G. S. Kinker, L. H. Ostrowski, P. A. C. Ribeiro, R. Chanoch, S. M. Muxel, I. Tirosh, G. Spadoni, S. Rivara, V. R. Martins, T. G. Santos, R. P. Markus, P. A. C. M. Fernandes, MT1 and MT2 melatonin receptors play opposite roles in brain cancer progression. *J. Mol. Med.* . **99**, 289–301 (2021).

203. M.-Y. Ou, X.-C. Ju, Y.-J. Cai, X.-Y. Sun, J.-F. Wang, X.-Q. Fu, Q. Sun, Z.-G. Luo, Heterogeneous nuclear ribonucleoprotein A3 controls mitotic progression of neural progenitors via interaction with cohesin. *Development*. **147**, dev185132 (2020).

204. M. Zech, F. Castrop, B. Schormair, A. Jochim, T. Wieland, N. Gross, P. Lichtner, A. Peters, C. Gieger, T. Meitinger, T. M. Strom, K. Oexle, B. Haslinger, J. Winkelmann, DYT16 revisited: exome sequencing identifies PRKRA mutations in a European dystonia family. *Mov. Disord.* **29**, 1504–1510 (2014).

205. J. Berciano, K. Peeters, A. García, T. López-Alburquerque, E. Gallardo, A. Hernández-Fabián, A. L. Pelayo-Negro, E. De Vriendt, J. Infante, A. Jordanova, NEFL N98S mutation: another cause of dominant intermediate Charcot-Marie-Tooth disease with heterogeneous early-onset phenotype. *J. Neurol.* **263**, 361–369 (2016).

206. J. Berciano, A. García, E. Gallardo, K. Peeters, A. L. Pelayo-Negro, S. Álvarez-Paradelo, J. Gazulla, M. Martínez-Tames, J. Infante, A. Jordanova, Intermediate Charcot-Marie-Tooth disease: an electrophysiological reappraisal and systematic review. *J. Neurol.* **264**, 1655–1677 (2017).

207. A. Jimenez-Pascual, J. S. Hale, A. Kordowski, J. Pugh, D. J. Silver, D. Bayik, G. Roversi, T. J. Alban, S. Rao, R. Chen, T. M. McIntyre, G. Colombo, G. Taraboletti, K. O. Holmberg, K. Forsberg-Nilsson, J. D. Lathia, F. A. Siebzehnrubl, ADAMDEC1 Maintains a Growth Factor Signaling Loop in Cancer Stem Cells. *Cancer Discov.* **9**, 1574–1589 (2019).

208. A. Badaloni, F. Casoni, L. Croci, F. Chiara, A. Bizzoca, G. Gennarini, O. Cremona, R. Hawkes, G. G. Consalez, Dynamic Expression and New Functions of Early B Cell Factor 2 in Cerebellar Development. *Cerebellum*. **18**, 999–1010 (2019).

209. M. Wagner, Y. Skorobogatko, B. Pode-Shakked, C. M. Powell, B. Alhaddad, A. Seibt, O. Barel, G. Heimer, C. Hoffmann, L. A. Demmer, Y. Perilla-Young, M. Remke, D. Wieczorek, T. Navaratnarajah, P. Lichtner, D. Klee, H. E. Shamseldin, F. Al Mutairi, E. Mayatepek, T. Strom, T. Meitinger, F. S. Alkuraya, Y. Anikster, A. R. Saltiel, F. Distelmaier, Bi-allelic Variants in RALGAPA1 Cause Profound Neurodevelopmental Disability, Muscular Hypotonia, Infantile

Spasms, and Feeding Abnormalities. *Am. J. Hum. Genet.* **106**, 246–255 (2020).

210. T. Guran, G. Yesil, S. Turan, Z. Atay, E. Bozkurtlar, A. Aghayev, S. Gul, I. Tinay, B. Aru, S. Arslan, M. K. Koroglu, F. Ercan, G. Y. Demirel, F. S. Eren, B. Karademir, A. Bereket, PPP2R3C gene variants cause syndromic 46,XY gonadal dysgenesis and impaired spermatogenesis in humans. *Eur. J. Endocrinol.* **180**, 291–309 (2019).

211. A. Carré, G. Szinnai, M. Castanet, S. Sura-Trueba, E. Tron, I. Broutin-L'Hermite, P. Barat, C. Goizet, D. Lacombe, M.-L. Moutard, C. Raybaud, C. Raynaud-Ravni, S. Romana, H. Ythier, J. Léger, M. Polak, Five new TTF1/NKX2.1 mutations in brain-lung-thyroid syndrome: rescue by PAX8 synergism in one case. *Hum. Mol. Genet.* **18**, 2266–2276 (2009).

212. W. Jarrar, J. M. Dias, J. Ericson, H.-H. Arnold, A. Holz, Nkx2.2 and Nkx2.9 are the key regulators to determine cell fate of branchial and visceral motor neurons in caudal hindbrain. *PLoS One*. **10**, e0124408 (2015).

213. V. Weitensteiner, R. Zhang, J. Bungenberg, M. Marks, J. Gehlen, D. J. Ralser, A. C. Hilger, A. Sharma, J. Schumacher, U. Gembruch, W. M. Merz, A. Becker, J. Altmüller, H. Thiele, B. G. Herrmann, B. Odermatt, M. Ludwig, H. Reutter, Exome sequencing in syndromic brain malformations identifies novel mutations in ACTB, and SLC9A6, and suggests BAZ1A as a new candidate gene. *Birth Defects Res*. **110**, 587–597 (2018).

214. E. Ponzi, M. Gentile, E. Agolini, E. Matera, R. Palumbi, A. L. Buonadonna, A. Peschechera, A. Gabellone, M. F. Antonucci, L. Margari, 14q12q13.2 microdeletion syndrome: Clinical characterization of a new patient, review of the literature, and further evidence of a candidate region for CNS anomalies. *Mol Genet Genomic Med*. **8**, e1289 (2020).

215. T. Suzuki, T. Suzuki, M. Raveau, N. Miyake, G. Sudo, Y. Tsurusaki, T. Watanabe, Y. Sugaya, T. Tatsukawa, E. Mazaki, A. Shimohata, I. Kushima, B. Aleksic, T. Shiino, T. Toyota, Y. Iwayama, K.

Nakaoka, I. Ohmori, A. Sasaki, K. Watanabe, S. Hirose, S. Kaneko, Y. Inoue, T. Yoshikawa, N. Ozaki, M. Kano, T. Shimoji, N. Matsumoto, K. Yamakawa, A recurrent PJA1 variant in trigonocephaly and neurodevelopmental disorders. *Ann Clin Transl Neurol*. **7**, 1117–1131 (2020).

216. S. R. F. Twigg, R. Kan, C. Babbs, E. G. Bochukova, S. P. Robertson, S. A. Wall, G. M. Morriss-Kay, A. O. M. Wilkie, Mutations of ephrin-B1 (EFNB1), a marker of tissue boundary formation, cause craniofrontonasal syndrome. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 8652–8657 (2004).

217. J. Y. Han, H. J. Kim, J. H. Jang, I. G. Lee, J. Park, Trio-Based Whole-Exome Sequencing Identifies a De novo EFNB1 Mutation as a Genetic Cause in Female Infant With Brain Anomaly and Developmental Delay. *Front Pediatr*. **8**, 461 (2020).

218. G. Zanni, Y. Saillour, M. Nagara, P. Billuart, L. Castelnau, C. Moraine, L. Faivre, E. Bertini, A. Durr, A. Guichet, D. Rodriguez, V. des Portes, C. Beldjord, J. Chelly, Oligophrenin 1 mutations frequently cause X-linked mental retardation with cerebellar hypoplasia. *Neurology*. **65**, 1364–1369 (2005).

219. A. Hoischen, B. W. M. van Bon, B. Rodríguez-Santiago, C. Gilissen, L. E. L. M. Vissers, P. de Vries, I. Janssen, B. van Lier, R. Hastings, S. F. Smithson, R. Newbury-Ecob, S. Kjaergaard, J. Goodship, R. McGowan, D. Bartholdi, A. Rauch, M. Peippo, J. M. Cobben, D. Wieczorek, G. Gillessen-Kaesbach, J. A. Veltman, H. G. Brunner, B. B. B. A. de Vries, De novo nonsense mutations in ASXL1 cause Bohring-Opitz syndrome. *Nat. Genet.* **43**, 729–731 (2011).

220. A. H. Alsabban, M. Morikawa, Y. Tanaka, Y. Takei, N. Hirokawa, Kinesin Kif3b mutation reduces NMDAR subunit NR2A trafficking and causes schizophrenia-like phenotypes in mice. *EMBO J.* **39**, e101090 (2020).

221. L. Adnani, R. Dixit, X. Chen, A. Balakrishnan, H. Modi, Y. Touahri, C. Logan, C. Schuurmans, Plag1 and Plagl2 have overlapping and distinct functions in telencephalic development. *Biol. Open*.

7, bio038661 (2018).

222. J. Wu, G. Lu, Multiple functions of TBCK protein in neurodevelopment disorders and tumors. *Oncol. Lett.* **21**, 17 (2021).

223. E. J. Bhoj, D. Li, M. Harr, S. Edvardson, O. Elpeleg, E. Chisholm, J. Juusola, G. Douglas, M. J. Guillen Sacoto, K. Siquier-Pernet, A. Saadi, C. Bole-Feysot, P. Nitschke, A. Narravula, M. Walke, M. B. Horner, D.-L. Day-Salvatore, P. Jayakar, S. A. S. Vergano, M. A. Tarnopolsky, M. Hegde, L. Colleaux, P. Crino, H. Hakonarson, Mutations in TBCK, Encoding TBC1-Domain-Containing Kinase, Lead to a Recognizable Syndrome of Intellectual Disability and Hypotonia. *Am. J. Hum. Genet.* **98**, 782–788 (2016).

224. T. Avşar, Ş. Çaliş, B. Yilmaz, G. Demirci Otluoğlu, C. Holyavkin, T. Kiliç, Genome-wide identification of Chiari malformation type I associated candidate genes and chromosomal variations. *Turk. J. Biol.* **44**, 449–456 (2020).

225. H. Behesti, T. R. Fore, P. Wu, Z. Horn, M. Leppert, C. Hull, M. E. Hatten, ASTN2 modulates synaptic strength by trafficking and degradation of surface proteins. *Proc. Natl. Acad. Sci. U. S. A.* **115**, E9717–E9726 (2018).

226. A. P. Chiang, J. S. Beck, H.-J. Yen, M. K. Tayeh, T. E. Scheetz, R. E. Swiderski, D. Y. Nishimura, T. A. Braun, K.-Y. A. Kim, J. Huang, K. Elbedour, R. Carmi, D. C. Slusarski, T. L. Casavant, E. M. Stone, V. C. Sheffield, Homozygosity mapping with SNP arrays identifies TRIM32, an E3 ubiquitin ligase, as a Bardet-Biedl syndrome gene (BBS11). *Proc. Natl. Acad. Sci. U. S. A.* **103**, 6287–6292 (2006).

227. V. Saccone, M. Palmieri, L. Passamano, G. Piluso, G. Meroni, L. Politano, V. Nigro, Mutations that impair interaction properties of TRIM32 associated with limb-girdle muscular dystrophy 2H. *Hum. Mutat.* **29**, 240–247 (2008).

228. C. J. Schoen, S. B. Emery, M. C. Thorne, H. R. Ammana, E. Sliwerska, J. Arnett, M. Hortsch, F. Hannan, M. Burmeister, M. M. Lesperance, Increased activity of Diaphanous homolog 3 (DIAPH3)/diaphanous causes hearing defects in humans with auditory neuropathy and in Drosophila. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 13396–13401 (2010).

229. E. O.-C. Lau, D. Damiani, G. Chehade, N. Ruiz-Reig, R. Saade, Y. Jossin, M. Aittaleb, O. Schakman, N. Tajeddine, P. Gailly, F. Tissir, DIAPH3 deficiency links microtubules to mitotic errors, defective neurogenesis, and brain dysfunction. *Elife*. **10**, e61974 (2021).

230. M. E. C. Meuwissen, R. Schot, S. Buta, G. Oudesluijs, S. Tinschert, S. D. Speer, Z. Li, L. van Unen, D. Heijsman, T. Goldmann, M. H. Lequin, J. M. Kros, W. Stam, M. Hermann, R. Willemsen, R. W. W. Brouwer, W. F. J. Van IJcken, M. Martin-Fernandez, I. de Coo, J. Dudink, F. A. T. de Vries, A. Bertoli Avella, M. Prinz, Y. J. Crow, F. W. Verheijen, S. Pellegrini, D. Bogunovic, G. M. S. Mancini, Human USP18 deficiency underlies type 1 interferonopathy leading to severe pseudo-TORCH syndrome. *J. Exp. Med.* **213**, 1163–1174 (2016).

231. A. T. Vulto-van Silfhout, T. Nakagawa, N. Bahi-Buisson, S. A. Haas, H. Hu, M. Bienek, L. E. L. M. Vissers, C. Gilissen, A. Tzschach, A. Busche, J. Müsebeck, P. Rump, I. B. Mathijssen, K. Avela, M. Somer, F. Doagu, A. K. Philips, A. Rauch, A. Baumer, K. Voesenek, K. Poirier, J. Vigneron, D. Amram, S. Odent, M. Nawara, E. Obersztyn, J. Lenart, A. Charzewska, N. Lebrun, U. Fischer, W. M. Nillesen, H. G. Yntema, I. Järvelä, H.-H. Ropers, B. B. A. de Vries, H. G. Brunner, H. van Bokhoven, F. L. Raymond, M. A. A. P. Willemsen, J. Chelly, Y. Xiong, A. J. Barkovich, V. M. Kalscheuer, T. Kleefstra, A. P. M. de Brouwer, Variants in CUL4B are associated with cerebral malformations. *Hum. Mutat.* **36**, 106–117 (2015).

232. M. Co, A. G. Anderson, G. Konopka, FOXP transcription factors in vertebrate brain development, function, and disorders. *Wiley Interdiscip. Rev. Dev. Biol.* **9**, e375 (2020).

233. H. Ito, R. Morishita, M. Mizuno, N. Kawamura, H. Tabata, K.-I. Nagata, Biochemical and Morphological Characterization of a Neurodevelopmental Disorder-Related Mono-ADP-Ribosylhydrolase, MACRO Domain Containing 2. *Dev. Neurosci.* **40**, 278-287 (2018).

234. H. Hor, L. Francescatto, L. Bartesaghi, S. Ortega-Cubero, M. Kousi, O. Lorenzo-Betancor, F. J. Jiménez-Jiménez, A. Gironell, J. Clarimón, O. Drechsel, J. A. G. Agúndez, D. Kenzelmann Broz, R. Chiquet-Ehrismann, A. Lleó, F. Coria, E. García-Martin, H. Alonso-Navarro, M. J. Martí, J. Kulisevsky, C. N. Hor, S. Ossowski, R. Chrast, N. Katsanis, P. Pastor, X. Estivill, Missense mutations in TENM4, a regulator of axon guidance and central myelination, cause essential tremor. *Hum. Mol. Genet.* **24**, 5677–5686 (2015).

235. N. Hamajima, M. Kouwaki, P. Vreken, K. Matsuda, S. Sumi, M. Imaeda, S. Ohba, K. Kidouchi, M. Nonaka, M. Sasaki, N. Tamaki, Y. Endo, R. De Abreu, J. Rotteveel, A. van Kuilenburg, A. van Gennip, H. Togari, Y. Wada, Dihydropyrimidinase deficiency: structural organization, chromosomal localization, and mutation analysis of the human dihydropyrimidinase gene. *Am. J. Hum. Genet.* **63**, 717–726 (1998).

236. S. Mechaussier, B. Almoallem, C. Zeitz, K. Van Schil, L. Jeddawi, J. Van Dorpe, A. Dueñas Rey, C. Condroyer, O. Pelle, M. Polak, N. Boddaert, N. Bahi-Buisson, M. Cavallin, J.-L. Bacquet, A. Mouallem-Bézière, O. Zambrowski, J. A. Sahel, I. Audo, J. Kaplan, J.-M. Rozet, E. De Baere, I. Perrault, Loss of Function of RIMS2 Causes a Syndromic Congenital Cone-Rod Synaptic Disease with Neurodevelopmental and Pancreatic Involvement. *Am. J. Hum. Genet.* **106**, 859–871 (2020).

237. P.-J. Mei, J. Bai, F.-A. Miao, C. Chen, Y.-S. Zhu, Z.-L. Li, J.-N. Zheng, Y.-C. Fan, CTHRC1 mediates multiple pathways regulating cell invasion, migration and adhesion in glioma. *Int. J. Clin. Exp. Pathol.* **10**, 9318–9329 (2017).

238. J. Nousbeck, R. Spiegel, A. Ishida-Yamamoto, M. Indelman, A. Shani-Adir, N. Adir, E. Lipkin,

S. Bercovici, D. Geiger, M. A. van Steensel, P. M. Steijlen, R. Bergman, A. Bindereif, M. Choder, S. Shalev, E. Sprecher, Alopecia, neurological defects, and endocrinopathy syndrome caused by decreased expression of RBM28, a nucleolar protein associated with ribosome biogenesis. *Am. J. Hum. Genet.* **82**, 1114–1121 (2008).

239. S. R. F. Twigg, R. B. Hufnagel, K. A. Miller, Y. Zhou, S. J. McGowan, J. Taylor, J. Craft, J. C. Taylor, S. L. Santoro, T. Huang, R. J. Hopkin, A. F. Brady, J. Clayton-Smith, C. L. Clericuzio, D. K. Grange, L. Groesser, C. Hafner, D. Horn, I. K. Temple, W. B. Dobyns, C. J. Curry, M. C. Jones, A. O. M. Wilkie, A Recurrent Mosaic Mutation in SMO, Encoding the Hedgehog Signal Transducer Smoothened, Is the Major Cause of Curry-Jones Syndrome. *Am. J. Hum. Genet.* **98**, 1256–1265 (2016).

240. T.-L. Le, Y. Sribudiani, X. Dong, C. Huber, C. Kois, G. Baujat, C. T. Gordon, V. Mayne, L. Galmiche, V. Serre, N. Goudin, M. Zarhrate, C. Bole-Feysot, C. Masson, P. Nitschké, F. W. Verheijen, L. Pais, A. Pelet, S. Sadedin, J. A. Pugh, N. Shur, S. M. White, S. El Chehadeh, J. Christodoulou, V. Cormier-Daire, R. M. W. Hofstra, S. Lyonnet, T. Y. Tan, T. Attié-Bitach, W. S. Kerstjens-Frederikse, J. Amiel, S. Thomas, Bi-allelic Variations of SMO in Humans Cause a Broad Spectrum of Developmental Anomalies Due to Abnormal Hedgehog Signaling. *Am. J. Hum. Genet.* **106**, 779–792 (2020).

241. O. V. Chechneva, F. Mayrhofer, D. J. Daugherty, R. G. Krishnamurty, P. Bannerman, D. E. Pleasure, W. Deng, A Smoothened receptor agonist is neuroprotective and promotes regeneration after ischemic brain injury. *Cell Death Dis.* **5**, e1481 (2014).

242. X. Fan, V. P. Masamsetti, J. Q. Sun, K. Engholm-Keller, P. Osteil, J. Studdert, M. E. Graham, N. Fossat, P. P. Tam, TWIST1 and chromatin regulatory proteins interact to guide neural crest cell differentiation. *Elife*. **10**, e62873 (2021).

243. M. Yu, L. Ma, Y. Yuan, X. Ye, A. Montagne, J. He, T.-V. Ho, Y. Wu, Z. Zhao, N. Sta Maria, R. Jacobs, M. Urata, H. Wang, B. V. Zlokovic, J.-F. Chen, Y. Chai, Cranial Suture Regeneration Mitigates Skull and Neurocognitive Defects in Craniosynostosis. *Cell*. **184**, 243–256.e18 (2021).

244. S. A. Mikheeva, A. M. Mikheev, A. Petit, R. Beyer, R. G. Oxford, L. Khorasani, J.-P. Maxwell, C. A. Glackin, H. Wakimoto, I. González-Herrero, I. Sánchez-García, J. R. Silber, P. J. Horner, R. C. Rostomily, TWIST1 promotes invasion through mesenchymal change in human glioblastoma. *Mol. Cancer*. **9**, 194 (2010).

245. S. Q. Shen, C. A. Myers, A. E. Hughes, L. C. Byrne, J. G. Flannery, J. C. Corbo, Massively parallel cis-regulatory analysis in the mammalian central nervous system. *Genome Res.* **26**, 238–255 (2016).

246. S. Zhang, J. Li, R. Lea, K. Vleminckx, E. Amaya, Fezf2 promotes neuronal differentiation through localised activation of Wnt/β-catenin signalling during forebrain development. *Development*. **141**, 4794–4805 (2014).

247. O. Mercati, A. Danckaert, G. André-Leroux, M. Bellinzoni, L. Gouder, K. Watanabe, Y. Shimoda, R. Grailhe, F. De Chaumont, T. Bourgeron, I. Cloëz-Tayarani, Contactin 4, -5 and -6 differentially regulate neuritogenesis while they display identical PTPRG binding sites. *Biol. Open*. **2**, 324–334 (2013).

248. C. Badouel, M. A. Zander, N. Liscio, M. Bagherie-Lachidan, R. Sopko, E. Coyaud, B. Raught, F. D. Miller, H. McNeill, Fat1 interacts with Fat4 to regulate neural tube closure, neural progenitor proliferation and apical constriction during mouse brain development. *Development*. **142**, 2781–2791 (2015).

249. Q. Zhang, X. Gao, C. Li, C. Feliciano, D. Wang, D. Zhou, Y. Mei, P. Monteiro, M. Anand, S. Itohara, X. Dong, Z. Fu, G. Feng, Impaired Dendritic Development and Memory in Sorbs2 Knock-

Out Mice. *J. Neurosci.* **36**, 2247–2260 (2016).

250. J. D. Ortigoza-Escobar, A. Oyarzabal, R. Montero, R. Artuch, C. Jou, C. Jiménez, L. Gort, P. Briones, J. Muchart, E. López-Gallardo, S. Emperador, E. R. Pesini, J. Montoya, B. Pérez, P. Rodríguez-Pombo, B. Pérez-Dueñas, Ndufs4 related Leigh syndrome: A case report and review of the literature. *Mitochondrion*. **28**, 73–78 (2016).

251. A. González-Quintana, M. J. Trujillo-Tiebas, A. L. Fernández-Perrone, A. Blázquez, A. Lucia, M. Morán, C. Ugalde, J. Arenas, C. Ayuso, M. A. Martín, Uniparental isodisomy as a cause of mitochondrial complex I respiratory chain disorder due to a novel splicing NDUFS4 mutation. *Mol. Genet. Metab.* **131**, 341–348 (2020).

252. E. Vantroys, A. Larson, M. Friederich, K. Knight, M. A. Swanson, C. A. Powell, J. Smet, S. Vergult, B. De Paepe, S. Seneca, H. Roeyers, B. Menten, M. Minczuk, A. Vanlander, J. Van Hove, R. Van Coster, New insights into the phenotype of FARS2 deficiency. *Mol. Genet. Metab.* **122**, 172–181 (2017).

253. R. Qin, S. Cao, T. Lyu, C. Qi, W. Zhang, Y. Wang, CDYL Deficiency Disrupts Neuronal Migration and Increases Susceptibility to Epilepsy. *Cell Rep.* **18**, 380–390 (2017).

254. A. Zito, D. Cartelli, G. Cappelletti, A. Cariboni, W. Andrews, J. Parnavelas, A. Poletti, M. Galbiati, Neuritin 1 promotes neuronal migration. *Brain Struct. Funct.* **219**, 105–118 (2014).

255. M. A. Kurian, E. Meyer, G. Vassallo, N. V. Morgan, N. Prakash, S. Pasha, N. A. Hai, S. Shuib, F. Rahman, E. Wassmer, J. H. Cross, F. J. O'Callaghan, J. P. Osborne, I. E. Scheffer, P. Gissen, E. R. Maher, Phospholipase C beta 1 deficiency is associated with early-onset epileptic encephalopathy. *Brain*. **133**, 2964–2970 (2010).

256. Y. R. Yang, D.-S. Kang, C. Lee, H. Seok, M. Y. Follo, L. Cocco, P.-G. Suh, Primary phospholipase C and brain disorders. *Adv. Biol. Regul.* **61**, 80–85 (2016).

257. R. Sueda, I. Imayoshi, Y. Harima, R. Kageyama, High Hes1 expression and resultant Ascl1 suppression regulate quiescent vs. active neural stem cells in the adult mouse brain. *Genes Dev.* **33**, 511–523 (2019).

258. T. Matsuzaki, T. Yoshihara, T. Ohtsuka, R. Kageyama, Hes1 expression in mature neurons in the adult mouse brain is required for normal behaviors. *Sci. Rep.* **9**, 8251 (2019).

259. L. El-Bazzal, A. Atkinson, A.-C. Gillart, M. Obeid, V. Delague, A. Mégarbané, A novel EXT2 mutation in a consanguineous family with severe developmental delay, microcephaly, seizures, feeding difficulties, and osteopenia extends the phenotypic spectrum of autosomal recessive EXT2-related syndrome (AREXT2). *Eur. J. Med. Genet.* **62**, 259–264 (2019).

260. A. Gupta, S. A. Ewing, D. L. Renaud, L. Hasadsri, K. M. Raymond, E. W. Klee, R. H. Gavrilova, Developmental delay, coarse facial features, and epilepsy in a patient with EXT2 gene variants. *Clin Case Rep*. **7**, 632–637 (2019).

261. O. B. Shevelev, V. I. Rykova, L. A. Fedoseeva, E. Y. Leberfarb, G. M. Dymshits, N. G. Kolosova, Expression of Ext1, Ext2, and heparanase genes in brain of senescence-accelerated OXYS rats in early ontogenesis and during development of neurodegenerative changes. *Biochemistry* . **77**, 56–61 (2012).

262. S. Hussain, Umm-E-Kalsoom, I. Ullah, K. Liaqat, S. Nawaz, W. Ahmad, A Novel Missense Variant in the ALX4 Gene Underlies Mild to Severe Frontonasal Dysplasia in a Consanguineous Family. *Genet. Test. Mol. Biomarkers*. **24**, 217–223 (2020).

263. M. Kojic, T. Gawda, M. Gaik, A. Begg, A. Salerno-Kochan, N. D. Kurniawan, A. Jones, K. Drożdżyk, A. Kościelniak, A. Chramiec-Głąbik, S. Hediyeh-Zadeh, M. Kasherman, W. J. Shim, E. Sinniah, L. A. Genovesi, R. K. Abrahamsen, C. D. Fenger, C. G. Madsen, J. S. Cohen, A. Fatemi, Z. Stark, S. Lunke, J. Lee, J. K. Hansen, M. F. Boxill, B. Keren, I. Marey, M. S. Saenz, K. Brown, S. A.

Alexander, S. Mureev, A. Batzilla, M. J. Davis, M. Piper, M. Bodén, T. H. J. Burne, N. J. Palpant, R. S. Møller, S. Glatt, B. J. Wainwright, Elp2 mutations perturb the epitranscriptome and lead to a complex neurodevelopmental phenotype. *Nat. Commun.* **12**, 2678 (2021).

264. Y.-T. Huang, J. O. Mason, D. J. Price, Lateral cortical Cdca7 expression levels are regulated by Pax6 and influence the production of intermediate progenitors. *BMC Neurosci.* **18**, 47 (2017).

265. Y. Wei, X. Han, C. Zhao, PDK1 regulates the survival of the developing cortical interneurons. *Mol. Brain*. **13**, 65 (2020).

Acknowledgements

We would like to thank D. Genereux, D. Levesque, K. Lindblad-Toh, and the members of the Pfenning Lab for useful discussions and suggestions. We would also like to thank P. Sullivan for curating the brain size residual annotations, A. Hindle for providing us with annotations of which mammals spend time underground, and M. Chikina, A. Kowalczyk, and E. Saputra for providing us early access to code for phylogenetic permulations for phylogenetic generalized least squares. 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 National Science Foundation grant number TG-BIO200055. Portions of this research were conducted on Lehigh University's Research Computing infrastructure, which is partially supported by NSF Award 2019035.

Funding: The Carnegie Mellon University Computational Biology Department Lane Fellowship supported I.M.K. An NIH NIDA DP1DA046585 grant supported I.M.K., A.J.L., D.E.S., M.E.W., C.S., X.Z., B.N.P., A.R.B., and A.R.P. The Alfred P. Sloan Foundation Research Fellowship supported I.M.K., M.E.W., and A.R.P. An NSF Graduate Research Fellowship Program under grants DGE1252522 and DGE1745016 supported A.J.L. A Carnegie Mellon University Summer Undergraduate Research Fellowship supported D.E.S. An NIH NIDA Fellowship grant F30DA053020 supported B.N.P.

Author contributions: I.M.K., A.J.L., and D.E.S. are listed as co-first authors in last name-alphabetical order because they contributed equally to the manuscript. Conceptualization was done by I.M.K. and A.R.P. Data curation was done by I.M.K. with assistance from C.S., B.N.P., A.J.L., W.K.M., and K.F. Formal analysis was done by I.M.K., D.E.S., and A.J.L. with assistance from C.S. and B.N.P. Funding acquisition was done by A.R.P. with assistance from A.J.L. and B.N.P. Investigation was done by I.M.K., A.J.L., D.E.S., and C.S. with assistance from M.E.W., B.N.P., K.P., A.R.B. and A.R.P. Methodology was developed by I.M.K., A.J.L., and D.E.S. with assistance from A.R.P. Supervision was done by I.M.K. and A.R.P. with assistance from A.J.L.,W.K.M., and M.E.W. Software was implemented by D.E.S, I.M.K., A.J.L., and C.S. with assistance from M.E.W., W.K.M., X.Z., and K.F. Visualization was done by I.M.K. and D.E.S. with assistance from C.S., A.J.L., and A.R.P. Manuscript preparation was done by I.M.K., D.E.S., and A.J.L. with assistance from A.R.P. and C.S. Manuscript review and editing was done by all authors.

Competing interests: No authors have any competing interests.

Data and materials availability: Publicly available ATAC-seq data was obtained from Gene Expression Omnibus accessions GSE161374, GSE146897, and GSE137311; China National GeneBank accession CNP0000198; and ArrayExpress accession E-MTAB-2633. Unpublished ATAC-seq data generated by the Pfenning Lab can be found at GSE159815 with token gzkbqwyobvmvfsd and will be released prior to publication. The tree used for the phenotype association pipeline can be obtained by contacting the Zoonomia Consortium and will be released prior to publication. Publicly available genomes and annotations were downloaded from NCBI Assembly and the UCSC Genome Browser. Publicly available Hi-C data was downloaded from http://hugin2.genetics.unc.edu/Project/hugin/. Motif discovery results and machine learning models can be found at http://daphne.compbio.cs.cmu.edu/files/ikaplow/TACITSupplement/. Machine learning model predictions can be obtained from authors upon request and will be released prior to publication. Code also used in our previous work (*25*) can be found at https://github.com/pfenninglab/OCROrthologPrediction, new code for this work can be found at https://github.com/pfenninglab/TACIT, and additional code can be obtained

from authors upon request.

ZOONOMIA CONSORITUM MEMBERS

Gregory Andrews[1], Joel C. Armstrong[2], Matteo Bianchi[3], Bruce W. Birren[4], Kevin R Bredemeyer[5], Ana M Breit[6], Matthew J Christmas[3], Joana Damas[7], Mark Diekhans[2], Michael X. Dong[3], Eduardo Eizirik[8], Kaili Fan[1], Cornelia Fanter[9], Nicole M. Foley[5], Karin Forsberg-Nilsson[10], Carlos J. Garcia[11], John Gatesy[12], Steven Gazal[13], Diane P. Genereux[4], Daniel Goodman[14], Linda Goodman[15], Jenna Grimshaw[11], Michaela K. Halsey[11], Andrew J Harris[5], Glenn Hickey[16], Michael Hiller[17], Allyson Hindle[9], Robert M. Hubley[18], Graham Hughes[19], Jeremy Johnson[4], David Juan[20], Irene M. Kaplow[21,22], Elinor K. Karlsson[1,4], Kathleen C. Keough[23,24], Bogdan Kirilenko[17], Jennifer M. Korstian[11], Sergey V. Kozyrev[3], Alyssa J. Lawler[25], Colleen Lawless[19], Danielle L. Levesque[6], Harris A. Lewin [7,26,27], Xue Li[1,4], Abigail Lind[23,24], Kerstin Lindblad-Toh[3,4], Voichita D. Marinescu[3], Tomas Marques-Bonet[20,28,29,30], Victor C Mason[31], Jennifer R. S. Meadows[3], Jill Moore[1], Diana D. Moreno-Santillan[11], Kathleen M. Morrill[1,4], Gerard Muntané[20], William J Murphy[5], Arcadi Navarro[20,32,33,34], Martin Nweeia[35,36,37,38], Austin Osmanski[11], Benedict Paten[2], Nicole S. Paulat[11], Eric Pederson[3], Andreas R. Pfenning[21,22], BaDoi N. Phan[21], Katherine S. Pollard[23,24,39], Kavya Prasad[21], Henry Pratt[1], David A. Ray[11], Jeb Rosen[18], Irina Ruf [40], Louise Ryan[19], Oliver A. Ryder[41,42], Daniel Schäffer[21], Aitor Serres[20], Beth Shapiro[43,44,] Arian F. A. Smit[18], Mark Springer[45], Chaitanya Srinivasan[21], Cynthia Steiner[46], Jessica M. Storer[18], Patrick F. Sullivan[47,48], Kevin A. M. Sullivan[10], Elisabeth Sundström[3], Megan A Supple[44], Ross Swofford[4], Joy-El Talbot[49], Emma Teeling[19], Jason Turner-Maier[4], Alejandro Valenzuela[20], Franziska Wagner[40], Ola Wallerman[3], Chao Wang[3], Juehan Wang[13], Zhiping Weng[1], Aryn P. Wilder[41], Morgan E. Wirthlin[21,22], Shuyang Yao[48], Xiaomeng Zhang[21]

1 Program in Bioinformatics and Integrative Biology, University of Massachusetts Medical School, Worcester, MA 01605, USA
2 Genomics Institute, UC Santa Cruz, 1156 High Street, Santa Cruz, CA 95064, USA
3 Science for Life Laboratory, Department of Medical Biochemistry and Microbiology, Uppsala University, Uppsala, 751 32, Sweden
4 Broad Institute of MIT and Harvard, Cambridge MA 02139, USA
5 Veterinary Integrative Biosciences, Texas A&M University, College Station, TX 77843, USA
6 School of Biology and Ecology, University of Maine, Orono, Maine 04469, USA
7 The Genome Center, University of California Davis, Davis, CA 95616, USA
8 School of Health and Life Sciences, Pontifical Catholic University of Rio Grande do Sul, Porto Alegre, 90619-900, Brazil
9 School of Life Sciences, University of Nevada Las Vegas, Las Vegas, NV 89154, USA
10 Department of Immunology, Genetics and Pathology, Science for Life Laboratory, Uppsala University, Uppsala, 751 85, Sweden
11 Department of Biological Sciences, Texas Tech University, Lubbock, TX 79409, USA
12 Division of Vertebrate Zoology, American Museum of Natural History, New York, NY 10024, USA
13 Keck School of Medicine, University of Southern California, Los Angeles, CA 90033, USA
14 University of California San Francisco, San Francisco, CA 94143 USA
15 Fauna Bio Inc., Emeryville, CA 94608, USA
16 Baskin School of Engineering, University of California Santa Cruz, Santa Cruz, CA 95064, USA
17 Max Planck Institute of Molecular Cell Biology and Genetics, 01307, Dresden, Germany
18 Institute for Systems Biology, Seattle, WA 98109, USA

19 School of Biology and Environmental Science, University College Dublin, Belfield, Dublin 4, Ireland
20 Institute of Evolutionary Biology (UPF-CSIC), Department of Experimental and Health Sciences, Universitat Pompeu Fabra, Barcelona, 08003, Spain
21 Department of Computational Biology, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA 15213, USA
22 Neuroscience Institute, Carnegie Mellon University, Pittsburgh, PA 15213, USA
23 Gladstone Institutes, San Francisco, CA 94158, USA
24 Department of Epidemiology & Biostatistics, University of California, San Francisco, CA 94158, USA
25 Department of Biology, Carnegie Mellon University, Pittsburgh, PA 15213, USA
26 Department of Evolution and Ecology, University of California, Davis, CA 95616, USA
27 John Muir Institute for the Environment, University of California, Davis, CA 95616, USA
28 Catalan Institution of Research and Advanced Studies (ICREA), 08010, Barcelona, Spain
29 CNAG-CRG, Centre for Genomic Regulation, Barcelona Institute of Science and Technology (BIST), 08036, Barcelona, Spain
30 Institut Català de Paleontologia Miquel Crusafont, Universitat Autònoma de Barcelona, 08193, Cerdanyola del Vallès, Barcelona, Spain
31 Institute of Cell Biology, University of Bern, 3012 Bern, Switzerland
32 Catalan Institution of Research and Advanced Studies (ICREA), 08010, Barcelona, Spain
33 CRG, Centre for Genomic Regulation, Barcelona Institute of Science and Technology (BIST), 08036, Barcelona, Spain
34 BarcelonaBeta Brain Research Center, Pasqual Maragall Foundation, Barcelona, 08005 Spain
35 Narwhal Genome Initiative, Department of Restorative Dentistry and Biomaterials Sciences, Harvard School of Dental Medicine, Boston, MA 02115, USA
36 Department of Comprehensive Care, School of Dental Medicine, Case Western Reserve University, Cleveland, OH 44106, USA
37 Department of Vertebrate Zoology, Smithsonian Institution, Washington, DC 20002, USA
38 Department of Vertebrate Zoology, Canadian Museum of Nature, Ottawa, Ontario K2P 2R1, Canada
39 Chan Zuckerberg Biohub, San Francisco, CA 94158, USA
40 Division of Messel Research and Mammalogy, Senckenberg Research Institute and Natural History Museum Frankfurt, 60325 Frankfurt am Main, Germany
41 Conservation Genetics, San Diego Zoo Wildlife Alliance, Escondido, CA 92027, USA
42 Department of Evolution, Behavior and Ecology, Division of Biology, University of California, San Diego, La Jolla, CA 92039 USA
43 Howard Hughes Medical Institute, University of California Santa Cruz, Santa Cruz, CA 95064, USA
44 Department of Ecology and Evolutionary Biology, University of California Santa Cruz, Santa Cruz, CA 95064, USA
45 Department of Evolution, Ecology and Organismal Biology, University of California, Riverside, CA 92521, USA
46 Conservation Science Wildlife Health, San Diego Zoo Wildlife Alliance, Escondido CA 92027, USA
47 Department of Genetics, University of North Carolina Medical School, Chapel Hill, NC 27599, USA
48 Department of Medical Epidemiology and Biostatistics, Karolinska Institute, Stockholm, Sweden
49 Iris Data Solutions, LLC, Orono, ME 04473, USA