

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

3

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

6

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

10 \*These authors contributed equally to this work.

11 **Affiliations:** Carnegie Mellon University Departments of <sup>1</sup>Computational Biology and <sup>2</sup>Biology and  
12 <sup>3</sup>Neuroscience Institute; <sup>4</sup>Current affiliation: Broad Institute, Massachusetts Institute of Technology;  
13 <sup>5</sup>Medical Scientist Training Program, University of Pittsburgh School of Medicine; <sup>6</sup>Department of  
14 Biological Sciences, Lehigh University; <sup>7</sup>Current affiliation: College of Law, University of Iowa

15 #Address correspondences to [apfenning@cmu.edu](mailto:apfenning@cmu.edu), [ikaplow@cs.cmu.edu](mailto:ikaplow@cs.cmu.edu)

16

17 **Abstract:**

18

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

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

30

31 **Main Text:**

32

33

34 INTRODUCTION

35

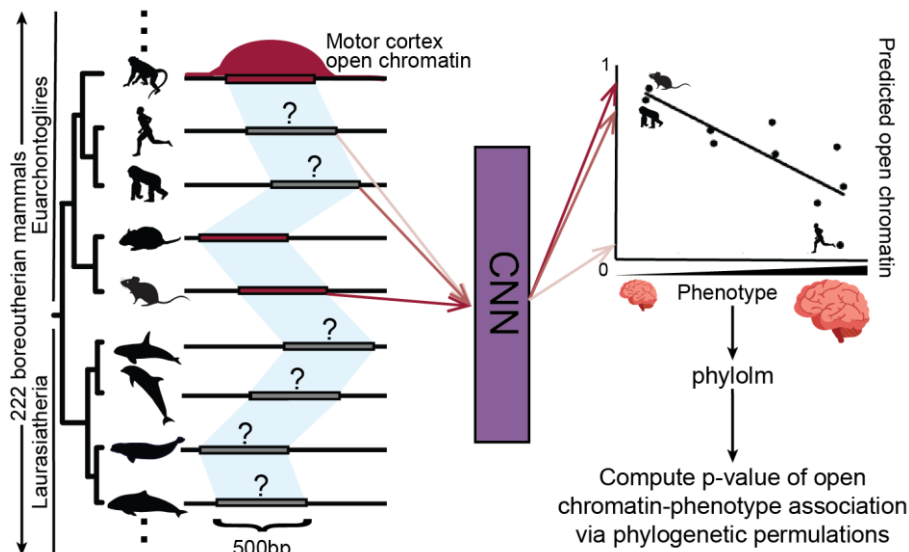
36 Much of the phenotypic diversity that exists across vertebrates is thought to have arisen from  
37 differences in how genes are expressed (1). Variation in phenotypes like vocal learning (2) and longevity  
38 (3) has been linked to patterns of gene expression within some of the most relevant brain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

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

53           Despite these successes, identifying enhancer-phenotype relationships is still a major challenge.  
54   Widely used methods to identify conservation and convergent evolution across orthologous genome  
55   sequences measure the extent to which the nucleotides within a given region align across species (17–19).  
56   While these approaches have led to some exciting findings (9, 20), many enhancer sequences and  
57   transcription factor binding sites are under less sequence constraint than promoter and gene sequences  
58   (21, 22). In fact, recent studies have shown that sequence conservation is not required for activity  
59   conservation at enhancer orthologs (23, 24) and can occur when enhancer activity is not conserved in a  
60   tissue of interest (25), so nucleotide sequence conservation at enhancers is sometimes an insufficient  
61   proxy for enhancer activity conservation.

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

79 chromatin data available from a relevant tissue or cell type in at least two species. It is therefore broadly  
80 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 Keeseey, 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.

81

82

## 83 RESULTS

84

### 85 Convolutional neural networks accurately predict open chromatin status of OCR orthologs

86 We applied TACIT to two tissues with open chromatin data from more than two species – motor

87 cortex and liver – as well as a tissue and a cell type with data from only two species – retina and motor

88 cortex PV+ interneurons. We used OCRs instead of other enhancer activity measures, such as H3K27ac

89 ChIP-seq regions, because OCRs tend to have a concentration of TF motifs near their summits and be

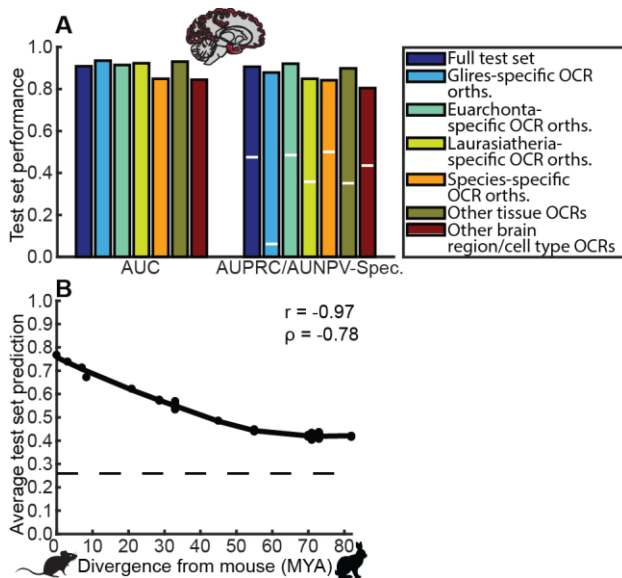
90 hundreds instead of thousands of base pairs long, allowing our model to focus on sequences likely to be

91 involved in enhancer activity and allowing us to easily map regions in species whose assemblies have

92 short scaffolds (14). We chose tissues and cell types that would demonstrate specificity in dissimilar  
93 tissues (brain versus liver) and have relationships with complex phenotypes of interest, including brain  
94 size, social behavior, and vocal learning. For tissues with more than two species, we trained CNNs to  
95 predict whether a region is an OCR or a non-OCR ortholog of an OCR, as described in our previous work  
96 (25).

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

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



**Figure 2: Motor cortex multi-species CNN performance.**

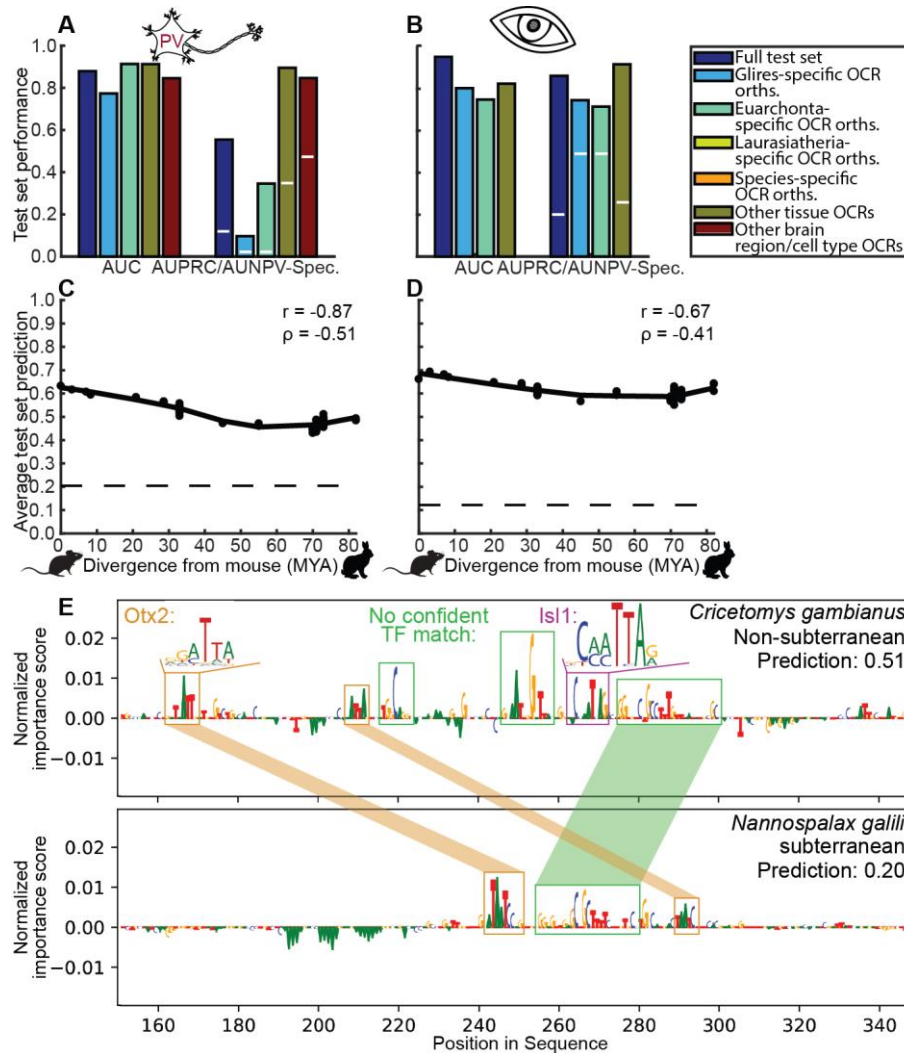
**A** shows the area under the ROC curve (AUROC) and the area under the precision-recall (if more negatives than positives)/negative predictive value-specificity (if more positives than negatives) curve (AUPRC/NPV-Spec.) for the full test set, clade-specific OCRs and non-OCRs, and shared versus tissue/brain 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.

120 Since no previous study has trained PV+ interneuron or retinal enhancer activity prediction  
121 models for predicting enhancer activity in species not used for training (25,27–29), we needed to  
122 investigate whether the PV+ interneuron and retinal regulatory codes are sufficiently conserved for  
123 accurately predicting open chromatin of OCR orthologs. We did this by running motif discovery on open  
124 chromatin datasets from each species for which data was available. For each of PV+ interneurons and  
125 retina, we found motifs for many of the same TFs in both species, and some of these TFs are known to be  
126 involved in PV+ interneurons and retina, respectively (**Supplementary Text, Supplementary Website**)  
127 (26).

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

131 large number of random regions of the genome with the same G/C-content as the positives with OCRs  
132 from other cell types or tissues, two negative sets that provided adequate performance for all of our  
133 metrics in our previous work (**Methods**) (25). To ensure that CNNs could make accurate predictions in  
134 species not used for training in our tissues and cell types, we first trained CNNs using only house mouse  
135 sequences and found that they achieved high lineage-specific OCR accuracy (AUC > 0.85 and  
136 AUPRC/NPV-Spec. > 0.60) as well as phylogeny-matching correlations (mean Pearson correlation < -  
137 0.65, mean Spearman correlation < -0.40 for retina and PV+ interneurons) for house mouse sequences  
138 (**Figs. S1B,C,E,F,H,I,K,L,N,O,Q,R, Tables S4-5**). The PV+ interneuron CNNs also achieved strong  
139 performance on human sequences (AUC > 0.70 and AUPRC/NPV-Spec. > fraction of examples in  
140 minority class for all criteria), where no human sequences were used in training as well as high tissue-  
141 specific OCR accuracy (AUC > 0.75 and AUPRC/NPV-Spec. > fraction of examples in minority class for  
142 all criteria), while the house mouse-trained retina CNNs did not work as well on human-specific OCRs  
143 and non-OCRs and liver non-retina OCRs. We then trained CNNs using sequences from both house  
144 mouse and human, and both the PV+ and retina CNNs achieved strong performance for all criteria (AUC  
145 > 0.70 and AUPRC/NPV-Spec. > fraction of examples in minority class for all criteria, mean Pearson  
146 correlation < -0.60, mean Spearman correlation < -0.40) (**Figs. 3A-D, Figs. S2B,C,E,F,H,I, Tables S6-**  
147 **7**).

148 To evaluate if our bulk tissue models were learning sequences relevant to the tissues in which  
149 they were trained, we interpreted what they had learned (**Methods**). Specifically, we computed the  
150 CNNs' per-nucleotide importance scores, which indicate the extent to which the CNN prioritizes the  
151 presence or absence of each nucleotide at each position (30, 31). We found that our CNNs seemed to have  
152 learned sequence patterns that are similar to motifs of TFs that are known to be involved in motor cortex  
153 and liver, such as MEF2C for motor cortex (32, 33) and HNF4A (34, 35) for liver, as well as sequence  
154 patterns that do not match any known TF motif (**Supplementary Text, Figs. S5-7**) (26). We then  
155 examined a specific retina OCR near the retina TF *Otx2*, where the OCR's orthologs in subterranean  
156 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.

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



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

167 From the four cross-species OCR datasets of interest (motor cortex, liver, PV+ interneuron, and  
168 retina), we identify 50,942,699 total orthologous regions across 222 Boreoeutherian mammals from  
169 402,880 total OCRs. Relative to human OCR annotations and phyloP annotations alone, we find that  
170 these predictions can provide a substantial boost for interpreting human disease-associated loci, with  
171 greater tissue- and cell type specificity. For example, in our other work, we found that human orthologs of  
172 regions predicted to have conserved motor cortex open chromatin are enriched for overlapping SNPs  
173 associated with schizophrenia, while human orthologs of regions predicted to have conserved liver open  
174 chromatin are enriched for overlapping SNPs associated with cholesterol-related traits (38, 39). These  
175 results demonstrate the power of TACIT to identify functionally relevant patterns of conservation.

176

## 177 **Applying TACIT to mammalian phenotypes**

178

179 *A framework for associating predicted open chromatin with phenotypes*

180

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

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

191  
192 *TACIT identifies motor cortex and PV+ interneuron OCRs associated with the evolution of brain size*  
193

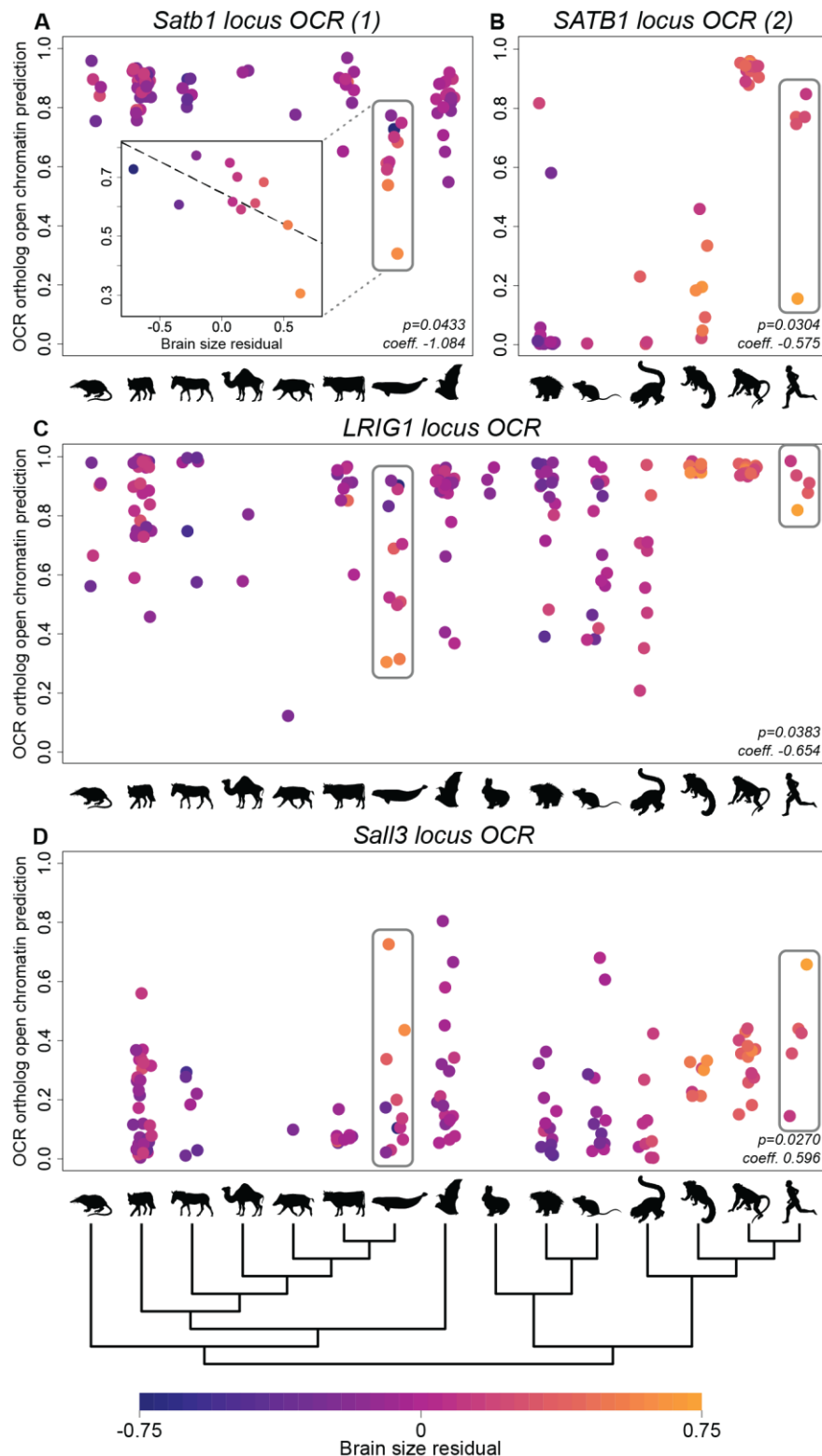
194 We used TACIT to identify motor cortex OCRs whose predicted open chromatin across mammals is  
195 significantly associated with brain size, a complex trait with great diversity across mammals that is  
196 thought to underlie human cognitive ability (42). As brain size scales with body size, we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

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

218 We identified two OCRs near *SATB1* — a gene with both microcephaly- and macrocephaly-  
219 associated mutations (48) — whose motor cortex predicted open chromatin status is significantly  
220 associated with brain size residual (**Fig. 4A-B, Figs. S8D,H**). For both of these associations, predicted  
221 open chromatin is associated with small brain size residual. The OCRs' coordinates in the genomes in  
222 which they were initially identified are chr17:52351209-52351928 (mm10) and chr2:174466184-  
223 174466517 (rheMac8). They are each about 500kb from the TSS of the gene, where one is upstream and  
224 the other is downstream. Neither OCR is near any other gene with a known connection to brain  
225 development; *Satb1/SATB1* is the second-closest gene to each, and the closer genes, *Kcnh8* and *TBC1D5*,  
226 each have known roles outside of brain growth (49, 50). The associations seem to be driven in large part  
227 by, respectively, cetaceans (**Fig. 4A**) and great apes (**Fig. 4B**), both of which have a large variation in  
228 brain size (51). In particular, the latter OCR is predicted to be active in all great apes except for humans,  
229 the great ape with the largest brain size residual. Interestingly, the reported case of *SATB1*-associated  
230 macrocephaly at birth was caused by a mutation that disrupts a large portion of the protein product, while  
231 microcephaly was usually reported with *SATB1* missense mutations (48). This pattern is consistent with  
232 the significant negative associations between predicted open chromatin and brain size residual, assuming  
233 that the OCRs we identified positively regulate the expression of *SATB1*.

234 We identified another OCR, chr2:75345159-75346046 (rheMac8), whose predicted motor cortex  
235 open chromatin also has a strong negative association with brain size residual in cetaceans (**Fig. 4C**). The  
236 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.**

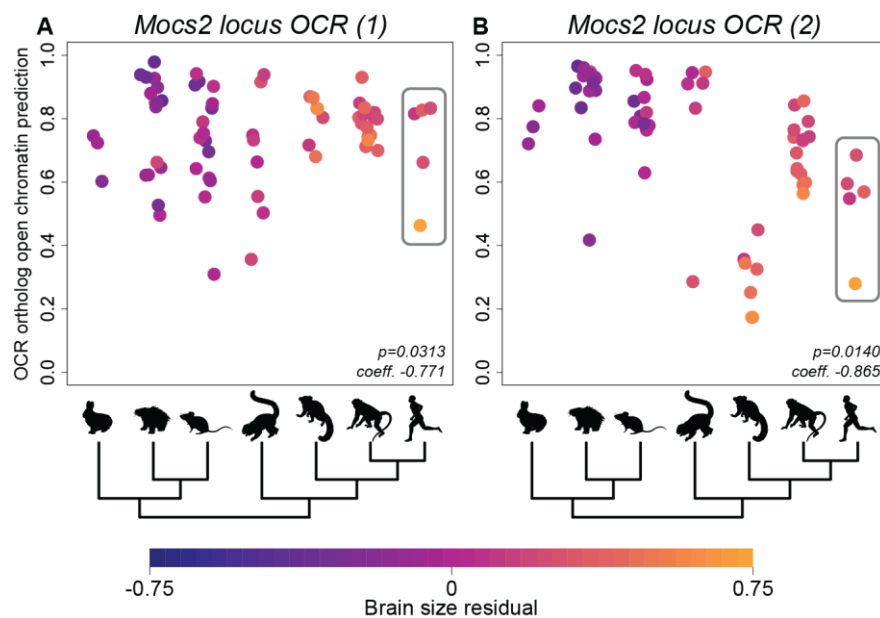
(A-B) highlight the negative association between predicted motor cortex open chromatin and brain size residual of two motor cortex OCRs in the *SATB1* locus, chr17:52351209-52351928 (mm10) and chr2:174466184-174466517 (rheMac8), within Laurasiatheria and Euarchontoglires, respectively. The latter OCR 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 permutations 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.

239 Also among the OCRs we identified near brain development genes is an OCR, chr18:81802310-  
240 81802951 (mm10), about 800kb from the gene *Sall3*. *Sall3* is the fourth-closest gene to this OCR, and  
241 one closer gene, *Mbp*, does have a connection to brain development (54). Hi-C from adult human cortex  
242 (55) shows that the bin containing the human ortholog of this OCR is close to *SALL3* in 3D space ( $p=2.3$   
243  $\times 10^{-11}$ , **Table S8**) but is not close to *MBP* ( $p=1$ ). This OCR displays a positive association with brain  
244 size residual both overall and within mammalian clades with especially large variations in brain size,  
245 including the great apes and cetaceans (**Fig. 4D**). *Sall3* is a member of the spalt-like family of  
246 transcription factors, which are important in development (56). Although a specific role of *Sall3* in motor  
247 cortex has not been described, there is evidence that *Sall3* regulates the maturation of neurons in other  
248 regions of the brain (57, 58), and *Sall3* is expressed in developing motor neurons (58) and human cerebral  
249 cortex (59).

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

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

265 114793737 (mm10) — are respectively about 60kb and 25kb from the *Mocs2* gene. Both have strong  
266 associations with brain size residual within Euarchonta (primates and their closest relatives), especially  
267 Hominoidea, and the first also has some association within Glires (rodents and their closest relatives)  
268 (Fig. 5A-B, respectively). *Mocs2* is one of four genes involved in Molybdenum cofactor biosynthesis  
269 (65). Molybdenum cofactor deficiency (MoCD) in humans is a rare, fatal disease marked by intractable  
270 seizures, hypoxia, and microcephaly (66). We also identified an OCR, chr1:95762160-95762660 (mm10),  
271 that is about 100kb away from the gene *St8sia4*, which is important for the development and density of  
272 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.

273 Interestingly, there is no overlap between the bulk motor cortex OCRs and PV+ interneuron  
274 OCRs with predicted activity that is significantly associated with brain size residual. In fact, no house  
275 mouse OCR ortholog from either set is within 5Mb of a house mouse OCR ortholog from the other set.  
276 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,

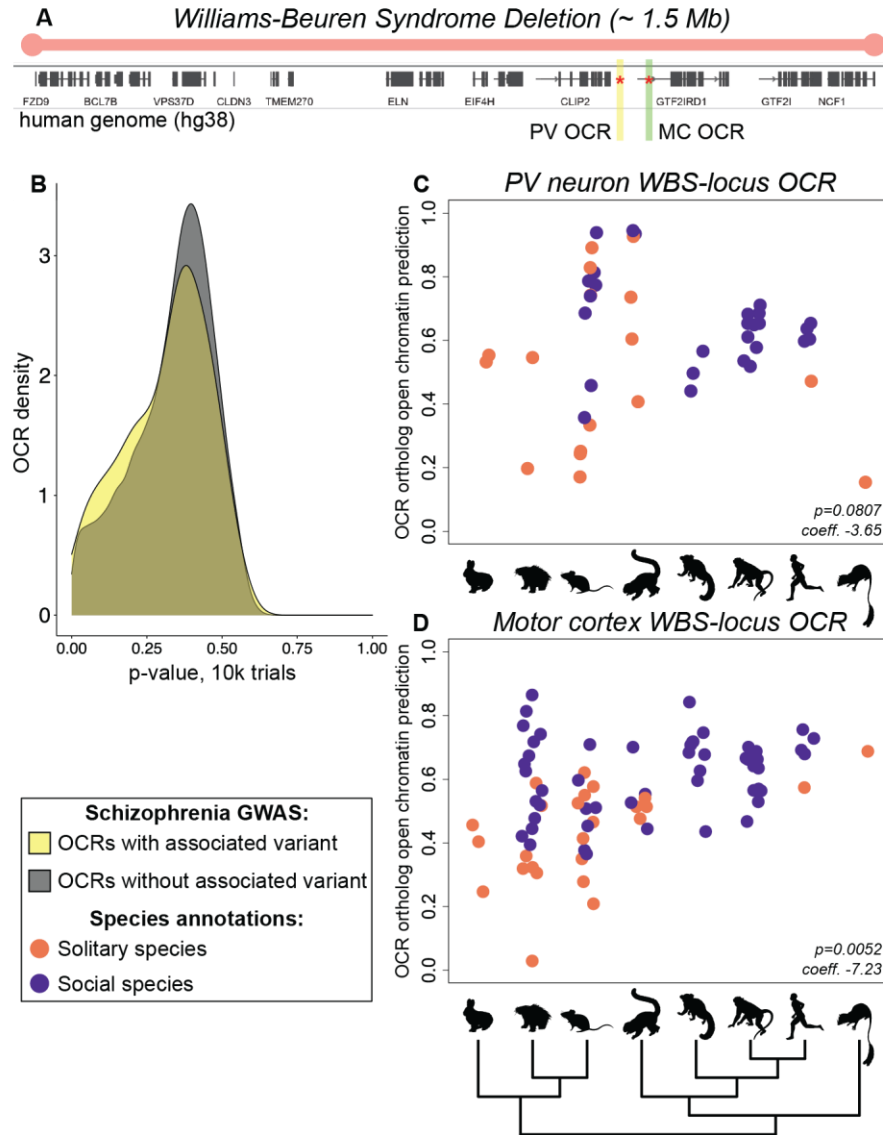
303 many extant lineages in disparate clades evolved towards social living strategies, including group living  
304 and breeding pair monogamy (76). Given the impact of PV+ neuron gene expression on social behaviors,  
305 we hypothesized that there might be PV+ OCR evolution associated with social structure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

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

315 When applying Cell-TACIT to only the WBS locus, we identified a mouse OCR (out of two  
316 OCRs in this locus) 29kb upstream of *GTF2IRD1* (human ortholog is 36kb upstream) that was marginally  
317 associated with non-solitary living ( $p = 0.08$ ) (**Fig. 6C**) and associated with group living ( $p = 0.02$ ). To  
318 evaluate whether this association was limited to PV+ interneurons, we also evaluated the relationship  
319 between predicted bulk motor cortex open chromatin and solitary as well as group living. For solitariness,  
320 we found one significantly negatively associated OCR ( $p = 0.005$ ) (**Fig. 6D**). This OCR is in an intron of  
321 *GTF2IRD1* that is about 26kb from its nearest TSS but does not overlap the OCR identified for PV+  
322 interneurons. For group living, we found two significantly associated OCRs, one of which is negatively  
323 associated ( $p = 0.04$ ) and the other of which is positively associated ( $p = 0.008$ ) and is the same OCR we  
324 found for solitariness. Of the 27 protein-coding genes in the WBS locus, *GTF2IRD1* is one of only two  
325 genes, where the other gene is its neighbor (*GTF2I*), with structural variants associated with the extreme  
326 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.

329

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

331

332 We applied TACIT and Cell-TACIT to vocal learning, the ability to modify vocal output as a result of  
333 social experience, which has convergently evolved across mammals and been associated with convergent  
334 patterns of gene expression in the motor cortex (2, 79, 80). We identified 42 OCRs displaying convergent  
335 patterns of predicted open chromatin after false discovery rate correction ( $\alpha=0.05$ ) for motor cortex tissue  
336 and 14 for PV+ interneurons, which are described in more depth in our other work . Notably, these vocal  
337 learning-associated OCRs showed some concordance with results obtained using complementary methods  
338 for detecting convergent evolution. One of the motor cortex OCRs lies 88kb from *Vip*, whose expression  
339 in the motor cortex has been associated with vocal learning (2). Another OCR is 715kb from *TSHZ3*,  
340 whose amino acid sequence also showed convergent evolution associated with vocal learning behavior ( $p$   
341  $< 0.0001$ , rank 3) (81). *TSHZ3* is involved in the formation of cortico-striatal circuits, which play a central  
342 role in vocal learning behavior in mammals and birds, and its disruption in the human population is  
343 associated with a form of autism that includes delayed or disrupted speech acquisition (80, 82).

344

345 DISCUSSION

346

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

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

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

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

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

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

420

## 421 LIST OF SUPPLEMENTARY MATERIALS

422 Materials and Methods

423 Supplementary Text

424 Supplementary Figures 1-8

425 Supplementary Tables 1-11

426

## 427 REFERENCES AND NOTES

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

- 429 (1975).
- 430 2. A. R. Pfenning, E. Hara, O. Whitney, M. V. Rivas, R. Wang, P. L. Roulhac, J. T. Howard, M.  
431 Wirthlin, P. V. Lovell, G. Ganapathy, J. Mouncastle, M. A. Moseley, J. W. Thompson, E. J.  
432 Soderblom, A. Iriki, M. Kato, M. T. Gilbert, G. Zhang, T. Bakken, A. Bongaarts, A. Bernard, E.  
433 Lein, C. V. Mello, A. J. Hartemink, E. D. Jarvis, Convergent transcriptional specializations in the  
434 brains of humans and song-learning birds. *Science*. **346**, 1256846 (2014).
- 435 3. A. A. Fushan, A. A. Turanov, S. G. Lee, E. B. Kim, A. V. Lobanov, S. H. Yim, R. Buffenstein, S. R.  
436 Lee, K. T. Chang, H. Rhee, J. S. Kim, K. S. Yang, V. N. Gladyshev, Gene expression defines natural  
437 changes in mammalian lifespan. *Aging Cell*. **14**, 352–365 (2015).
- 438 4. G. A. Wray, The evolutionary significance of cis-regulatory mutations. *Nat. Rev. Genet.* **8**, 206–216  
439 (2007).
- 440 5. D. Villar, P. Flicek, D. T. Odom, Evolution of transcription factor binding in metazoans —  
441 mechanisms and functional implications. *Nat. Rev. Genet.* **15**, 221–233 (2014).
- 442 6. E. Z. Kvon, O. K. Kamneva, U. S. Melo, I. Barozzi, M. Osterwalder, B. J. Mannion, V. Tissières, C.  
443 S. Pickle, I. Plajzer-Frick, E. A. Lee, M. Kato, T. H. Garvin, J. A. Akiyama, V. Afzal, J. Lopez-Rios,  
444 E. M. Rubin, D. E. Dickel, L. A. Pennacchio, A. Visel, Progressive Loss of Function in a Limb  
445 Enhancer during Snake Evolution. *Cell*. **167**, 633–642.e11 (2016).
- 446 7. L. A. Lettice, A. E. Hill, P. S. Devenney, R. E. Hill, Point mutations in a distant sonic hedgehog cis-  
447 regulator generate a variable regulatory output responsible for preaxial polydactyly. *Hum. Mol.*  
448 *Genet.* **17**, 978–985 (2008).
- 449 8. D. Furniss, L. A. Lettice, I. B. Taylor, P. S. Critchley, H. Giele, R. E. Hill, A. O. M. Wilkie, A  
450 variant in the sonic hedgehog regulatory sequence (ZRS) is associated with triphalangeal thumb and  
451 deregulates expression in the developing limb. *Hum. Mol. Genet.* **17**, 2417–2423 (2008).

- 452 9. R. Partha, B. K. Chauhan, Z. Ferreira, J. D. Robinson, K. Lathrop, K. K. Nischal, M. Chikina, N. L.  
453 Clark, Subterranean mammals show convergent regression in ocular genes and enhancers, along with  
454 adaptation to tunneling. *Elife*. **6**, e25884 (2017).
- 455 10. C. Y. McLean, P. L. Reno, A. A. Pollen, A. I. Bassan, T. D. Capellini, C. Guenther, V. B. Indjeian,  
456 X. Lim, D. B. Menke, B. T. Schaar, A. M. Wenger, G. Bejerano, D. M. Kingsley, Human-specific  
457 loss of regulatory DNA and the evolution of human-specific traits. *Nature*. **471**, 216–219 (2011).
- 458 11. Zoonomia Consortium, A comparative genomics multitool for scientific discovery and conservation.  
459 *Nature*. **587**, 240–245 (2020).
- 460 12. J. Armstrong, G. Hickey, M. Diekhans, I. T. Fiddes, A. M. Novak, A. Deran, Q. Fang, D. Xie, S.  
461 Feng, J. Stiller, D. Genreux, J. Johnson, V. D. Marinescu, J. Alföldi, R. S. Harris, K. Lindblad-Toh,  
462 D. Haussler, E. Karlsson, E. D. Jarvis, G. Zhang, B. Paten, Progressive Cactus is a multiple-genome  
463 aligner for the thousand-genome era. *Nature*. **587**, 246–251 (2020).
- 464 13. G. Hickey, B. Paten, D. Earl, D. Zerbino, D. Haussler, HAL: a hierarchical format for storing and  
465 analyzing multiple genome alignments. *Bioinformatics*. **29**, 1341–1342 (2013).
- 466 14. X. Zhang, I. M. Kaplow, M. Wirthlin, T. Y. Park, A. R. Pfenning, HALPER facilitates the  
467 identification of regulatory element orthologs across species. *Bioinformatics*. **36**, 4339–4340 (2020).
- 468 15. L. S. T. Ho, C. Ané, A linear-time algorithm for Gaussian and non-Gaussian trait evolution models.  
469 *Syst. Biol.* **63**, 397–408 (2014).
- 470 16. E. Saputra, A. Kowalczyk, L. Cusick, N. Clark, M. Chikina, Phylogenetic Permutations: A  
471 Statistically Rigorous Approach to Measure Confidence in Associations in a Phylogenetic Context.  
472 *Mol. Biol. Evol.* **38**, 3004–3021 (2021).
- 473 17. K. S. Pollard, M. J. Hubisz, K. R. Rosenbloom, A. Siepel, Detection of nonneutral substitution rates

- 474 on mammalian phylogenies. *Genome Res.* **20**, 110–121 (2010).
- 475 18. Z. Yang, PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**, 1586–1591  
476 (2007).
- 477 19. A. Kowalczyk, W. K. Meyer, R. Partha, W. Mao, N. L. Clark, M. Chikina, RERconverge: an R  
478 package for associating evolutionary rates with convergent traits. *Bioinformatics.* **35**, 4815–4817  
479 (2019).
- 480 20. J. G. Roscito, K. Sameith, G. Parra, B. E. Langer, A. Petzold, C. Moebius, M. Bickle, M. T.  
481 Rodrigues, M. Hiller, Phenotype loss is associated with widespread divergence of the gene  
482 regulatory landscape in evolution. *Nat. Commun.* **9**, 4737 (2018).
- 483 21. D. Villar, C. Berthelot, S. Aldridge, T. F. Rayner, M. Lukk, M. Pignatelli, T. J. Park, R. Deaville, J.  
484 T. Erichsen, A. J. Jasinska, J. M. A. Turner, M. F. Bertelsen, E. P. Murchison, P. Flicek, D. T.  
485 Odom, Enhancer Evolution across 20 Mammalian Species. *Cell.* **160**, 554–566 (2015).
- 486 22. K. Lindblad-Toh, M. Garber, O. Zuk, M. F. Lin, B. J. Parker, S. Washietl, P. Kheradpour, J. Ernst,  
487 G. Jordan, E. Mauceli, L. D. Ward, C. B. Lowe, A. K. Holloway, M. Clamp, S. Gnerre, J. Alföldi, K.  
488 Beal, J. Chang, H. Clawson, J. Cuff, F. Di Palma, S. Fitzgerald, P. Flicek, M. Guttman, M. J. Hubisz,  
489 D. B. Jaffe, I. Jungreis, W. J. Kent, D. Kostka, M. Lara, A. L. Martins, T. Massingham, I. Moltke, B.  
490 J. Raney, M. D. Rasmussen, J. Robinson, A. Stark, A. J. Vilella, J. Wen, X. Xie, M. C. Zody, Broad  
491 Institute Sequencing Platform and Whole Genome Assembly Team, J. Baldwin, T. Bloom, C. W.  
492 Chin, D. Heiman, R. Nicol, C. Nusbaum, S. Young, J. Wilkinson, K. C. Worley, C. L. Kovar, D. M.  
493 Muzny, R. A. Gibbs, Baylor College of Medicine Human Genome Sequencing Center Sequencing  
494 Team, A. Cree, H. H. Dihn, G. Fowler, S. Jhangiani, V. Joshi, S. Lee, L. R. Lewis, L. V. Nazareth,  
495 G. Okwuonu, J. Santibanez, W. C. Warren, E. R. Mardis, G. M. Weinstock, R. K. Wilson, Genome  
496 Institute at Washington University, K. Delehaunty, D. Dooling, C. Fronik, L. Fulton, B. Fulton, T.



- 497 Graves, P. Minx, E. Sodergren, E. Birney, E. H. Margulies, J. Herrero, E. D. Green, D. Haussler, A.  
498 Siepel, N. Goldman, K. S. Pollard, J. S. Pedersen, E. S. Lander, M. Kellis, A high-resolution map of  
499 human evolutionary constraint using 29 mammals. *Nature*. **478**, 476–482 (2011).
- 500 23. V. Snetkova, A. R. Ypsilanti, J. A. Akiyama, B. J. Mannion, I. Plajzer-Frick, C. S. Novak, A. N.  
501 Harrington, Q. T. Pham, M. Kato, Y. Zhu, J. Godoy, E. Meky, R. D. Hunter, M. Shi, E. Z. Kvon, V.  
502 Afzal, S. Tran, J. L. R. Rubenstein, A. Visel, L. A. Pennacchio, D. E. Dickel, Ultraconserved  
503 enhancer function does not require perfect sequence conservation. *Nat. Genet.* **53**, 521–528 (2021).
- 504 24. E. S. Wong, D. Zheng, S. Z. Tan, N. L. Bower, V. Garside, G. Vanwalleghem, F. Gaiti, E. Scott, B.  
505 M. Hogan, K. Kikuchi, E. McGlinn, M. Francois, B. M. Degnan, Deep conservation of the enhancer  
506 regulatory code in animals. *Science*. **370**, eaax8137 (2020).
- 507 25. I. M. Kaplow, D. E. Schäffer, M. E. Wirthlin, A. J. Lawler, A. R. Brown, M. Kleyman, A. R.  
508 Pfenning, <https://www.biorxiv.org/content/10.1101/2020.12.04.410795v3>.
- 509 26. Materials and methods are available as supplementary materials at the Science website.
- 510 27. L. Chen, A. E. Fish, J. A. Capra, Prediction of gene regulatory enhancers across species reveals  
511 evolutionarily conserved sequence properties. *PLoS Comput. Biol.* **14**, e1006484 (2018).
- 512 28. D. R. Kelley, Cross-species regulatory sequence activity prediction. *PLoS Comput. Biol.* **16**,  
513 e1008050 (2020).
- 514 29. L. Minnoye, I. I. Taskiran, D. Mauduit, M. Fazio, L. Van Aerschot, G. Hulselmans, V. Christiaens,  
515 S. Makhzami, M. Seltenhammer, P. Karras, A. Primot, E. Cadieu, E. van Rooijen, J.-C. Marine, G.  
516 Egidy, G.-E. Ghanem, L. Zon, J. Wouters, S. Aerts, Cross-species analysis of enhancer logic using  
517 deep learning. *Genome Res.* **30**, 1815–1834 (2020).
- 518 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,

- 541 Zoonomia Project Consortium, A. R. Pfenning, Fine-mapping candidate neurological disease-  
542 associated regulatory variants using cell type-aware comparative genomics. *Submitted to Science*  
543 *with Zoonomia Consortium*.
- 544 39. P. Sullivan, Zoonomia Consortium, Evolutionary constraint uniquely informs human disease  
545 mechanism. *Submitted to Science with Zoonomia Consortium*.
- 546 40. A. Grafen, The phylogenetic regression. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **326**, 119–157  
547 (1989).
- 548 41. A. R. Ives, T. Garland Jr, Phylogenetic logistic regression for binary dependent variables. *Syst. Biol.*  
549 **59**, 9–26 (2010).
- 550 42. C. Mitchell, D. L. Silver, Enhancing our brains: Genomic mechanisms underlying cortical evolution.  
551 *Semin. Cell Dev. Biol.* **76**, 23–32 (2018).
- 552 43. J. R. Burger, M. A. George, C. Leadbetter, F. Shaikh, The allometry of brain size in mammals. *J.*  
553 *Mammal.* **100**, 276–283 (2019).
- 554 44. S. Herculano-Houzel, The remarkable, yet not extraordinary, human brain as a scaled-up primate  
555 brain and its associated cost. *Proc. Natl. Acad. Sci. U. S. A.* **109 Suppl 1**, 10661–10668 (2012).
- 556 45. Y. Benjamini, Y. Hochberg, Controlling the false discovery rate: A practical and powerful approach  
557 to multiple testing. *J. R. Stat. Soc. Series B Stat. Methodol.* **57**, 289–300 (1995).
- 558 46. J. Liu, D. L. Silver, Founder cells shape brain evolution. *Cell.* **184**, 1965–1967 (2021).
- 559 47. McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University (Baltimore, MD),  
560 Online Mendelian Inheritance in Man, OMIM®. <https://omim.org/>.
- 561 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. Cimbaliestienè, H. Coon, DDD Study, E. C. Délot, F. Démurger,  
564 A.-S. Denommé-Pichon, C. Depienne, D. Donnai, D. A. Dymont, 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<sup>+</sup> 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.

- 586 D. Miller, S. A. Yuzwa, LRIG1-Mediated Inhibition of EGF Receptor Signaling Regulates Neural  
587 Precursor Cell Proliferation in the Neocortex. *Cell Rep.* **33**, 108257 (2020).
- 588 53. M. Á. Marqués-Torrejón, C. A. C. Williams, B. Southgate, N. Alfazema, M. P. Clements, C. Garcia-  
589 Diaz, C. Blin, N. Arranz-Emparan, J. Fraser, N. Gammoh, S. Parrinello, S. M. Pollard, LRIG1 is a  
590 gatekeeper to exit from quiescence in adult neural stem cells. *Nat. Commun.* **12**, 2594 (2021).
- 591 54. C. M. Deber, S. J. Reynolds, Central nervous system myelin: structure, function, and pathology.  
592 *Clin. Biochem.* **24**, 113–134 (1991).
- 593 55. P. Giusti-Rodríguez, L. Lu, Y. Yang, C. A. Crowley, X. Liu, I. Juric, J. S. Martin, A. Abnoui, S.  
594 Colby Allred, N. Ancalade, N. J. Bray, G. Breen, J. Bryois, C. M. Bulik, J. J. Crowley, J.  
595 Guintivano, P. R. Jansen, G. J. Jurjus, Y. Li, G. Mahajan, S. Marzi, J. Mill, M. C. O’Donovan, J. C.  
596 Overholser, M. J. Owen, A. F. Pardiñas, S. Pochareddy, D. Posthuma, G. Rajkowska, G. Santpere, J.  
597 E. Savage, N. Sestan, Y. Shin, C. A. Stockmeier, J. T. R. Walters, S. Yao, Bipolar Disorder Working  
598 Group of the Psychiatric Genomics Consortium, Eating Disorders Working Group of the Psychiatric  
599 Genomics Consortium, G. E. Crawford, F. Jin, M. Hu, Y. Li, P. F. Sullivan,  
600 <https://www.biorxiv.org/content/10.1101/406330v2>.
- 601 56. J. F. de Celis, R. Barrio, Regulation and function of Spalt proteins during animal development. *Int. J.*  
602 *Dev. Biol.* **53**, 1385–1398 (2009).
- 603 57. S. J. Harrison, M. Parrish, A. P. Monaghan, Sall3 is required for the terminal maturation of olfactory  
604 glomerular interneurons. *J. Comp. Neurol.* **507**, 1780–1794 (2008).
- 605 58. M. Parrish, T. Ott, C. Lance-Jones, G. Schuetz, A. Schwaeger-Nickolenko, A. P. Monaghan, Loss of  
606 the Sall3 gene leads to palate deficiency, abnormalities in cranial nerves, and perinatal lethality. *Mol.*  
607 *Cell. Biol.* **24**, 7102–7112 (2004).
- 608 59. M. Uhlén, L. Fagerberg, B. M. Hallström, C. Lindskog, P. Oksvold, A. Mardinoglu, Å. Sivertsson,

- 609 C. Kampf, E. Sjöstedt, A. Asplund, I. Olsson, K. Edlund, E. Lundberg, S. Navani, C. A.-K.  
610 Szigyarto, J. Odeberg, D. Djureinovic, J. O. Takanen, S. Hober, T. Alm, P.-H. Edqvist, H. Berling,  
611 H. Tegel, J. Mulder, J. Rockberg, P. Nilsson, J. M. Schwenk, M. Hamsten, K. von Feilitzen, M.  
612 Forsberg, L. Persson, F. Johansson, M. Zwahlen, G. von Heijne, J. Nielsen, F. Pontén, Proteomics.  
613 Tissue-based map of the human proteome. *Science*. **347**, 1260419 (2015).
- 614 60. A. J. Lawler, E. Ramamurthy, A. R. Brown, N. Shin, Y. Kim, N. Toong, I. M. Kaplow, M. Wirthlin,  
615 X. Zhang, G. Fox, A. R. Pfenning, <https://www.biorxiv.org/content/10.1101/2021.04.15.439984v1>.
- 616 61. A. Mo, E. A. Mukamel, F. P. Davis, C. Luo, G. L. Henry, S. Picard, M. A. Urich, J. R. Nery, T. J.  
617 Sejnowski, R. Lister, S. R. Eddy, J. R. Ecker, J. Nathans, Epigenomic Signatures of Neuronal  
618 Diversity in the Mammalian Brain. *Neuron*. **86**, 1369–1384 (2015).
- 619 62. B. Rudy, G. Fishell, S. Lee, J. Hjerling-Leffler, Three groups of interneurons account for nearly  
620 100% of neocortical GABAergic neurons. *Dev. Neurobiol.* **71**, 45–61 (2011).
- 621 63. P. McColgan, J. Joubert, S. J. Tabrizi, G. Rees, The human motor cortex microcircuit: insights for  
622 neurodegenerative disease. *Nat. Rev. Neurosci.* **21**, 401–415 (2020).
- 623 64. G. Gonzalez-Burgos, R. Y. Cho, D. A. Lewis, Alterations in cortical network oscillations and  
624 parvalbumin neurons in schizophrenia. *Biol. Psychiatry*. **77**, 1031–1040 (2015).
- 625 65. S. Leimkühler, A. Freuer, J. A. S. Araujo, K. V. Rajagopalan, R. R. Mendel, Mechanistic studies of  
626 human molybdopterin synthase reaction and characterization of mutants identified in group B  
627 patients of molybdenum cofactor deficiency. *J. Biol. Chem.* **278**, 26127–26134 (2003).
- 628 66. E. Bayram, Y. Topcu, P. Karakaya, U. Yis, H. Cakmakci, K. Ichida, S. H. Kurul, Molybdenum  
629 cofactor deficiency: Review of 12 cases (MoCD and review). *Eur. J. Paediatr. Neurol.* **17**, 1–6  
630 (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. Yartsev, A. P. Pfenning, Vocal learning-associated convergent evolution across mammals in  
674 regulatory element and protein sequence. *Submitted to Science with Zoonomia Consortium.*
- 675 82. D. Chabbert, X. Caubit, P. L. Roubertoux, M. Carlier, B. Habermann, B. Jacq, P. Salin, M. Metwaly,



- 676 C. Frahm, A. Fatmi, A. N. Garratt, D. Severac, E. Dubois, L. Kerkerian-Le Goff, L. Fasano, P.  
677 Gubellini, Postnatal Tshz3 Deletion Drives Altered Corticostriatal Function and Autism Spectrum  
678 Disorder-like Behavior. *Biol. Psychiatry*. **86**, 274–285 (2019).
- 679 83. S. S. P. Rao, M. H. Huntley, N. C. Durand, E. K. Stamenova, I. D. Bochkov, J. T. Robinson, A. L.  
680 Sanborn, I. Machol, A. D. Omer, E. S. Lander, E. L. Aiden, A 3D Map of the Human Genome at  
681 Kilobase Resolution Reveals Principles of Chromatin Looping. *Cell*. **159**, 1665–1680 (2014).
- 682 84. S. Volland, J. Esteve-Rudd, J. Hoo, C. Yee, D. S. Williams, A comparison of some organizational  
683 characteristics of the mouse central retina and the human macula. *PLoS One*. **10**, e0125631 (2015).
- 684 85. Y. Wu, H. Wang, H. Wang, E. A. Hadly, Rethinking the Origin of Primates by Reconstructing Their  
685 Diel Activity Patterns Using Genetics and Morphology. *Sci. Rep.* **7**, 11837 (2017).
- 686 86. M. T. Weirauch, A. Yang, M. Albu, A. G. Cote, A. Montenegro-Montero, P. Drewe, H. S.  
687 Najafabadi, S. A. Lambert, I. Mann, K. Cook, H. Zheng, A. Goity, H. van Bakel, J.-C. Lozano, M.  
688 Galli, M. G. Lewsey, E. Huang, T. Mukherjee, X. Chen, J. S. Reece-Hoyes, S. Govindarajan, G.  
689 Shaulsky, A. J. M. Walhout, F.-Y. Bouget, G. Ratsch, L. F. Larrondo, J. R. Ecker, T. R. Hughes,  
690 Determination and inference of eukaryotic transcription factor sequence specificity. *Cell*. **158**, 1431–  
691 1443 (2014).
- 692 87. T. L. Bailey, M. Boden, F. A. Buske, M. Frith, C. E. Grant, L. Clementi, J. Ren, W. W. Li, W. S.  
693 Noble, MEME Suite: Tools for motif discovery and searching. *Nucleic Acids Res.* **37**, W202–W208  
694 (2009).
- 695 88. C. Srinivasan, B. N. Phan, A. J. Lawler, E. Ramamurthy, M. Kleyman, A. R. Brown, I. M. Kaplow,  
696 M. E. Wirthlin, A. R. Pfenning, Addiction-associated genetic variants implicate brain cell type- and  
697 region-specific cis-regulatory elements in addiction neurobiology. *J. Neurosci.* JN-RM-2534-20  
698 (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. Ghasvand, 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.

722 M. Krienen, S. A. Sorensen, J. Eggermont, Z. Yao, B. D. Aevermann, A. I. Aldridge, A. Bartlett, D.  
723 Bertagnolli, T. Casper, R. G. Castanon, K. Crichton, T. L. Daigle, R. Dalley, N. Dee, N. Dembrow,  
724 D. Diep, S.-L. Ding, W. Dong, R. Fang, S. Fischer, M. Goldman, J. Goldy, L. T. Graybuck, B. R.  
725 Herb, X. Hou, J. Kancherla, M. Kroll, K. Lathia, B. van Lew, Y. E. Li, C. S. Liu, H. Liu, J. D.  
726 Lucero, A. Mahurkar, D. McMillen, J. A. Miller, M. Moussa, J. R. Nery, P. R. Nicovich, S.-Y. Niu,  
727 J. Orvis, J. K. Osteen, S. Owen, C. R. Palmer, T. Pham, N. Plongthongkum, O. Poirion, N. M. Reed,  
728 C. Rimorin, A. Rivkin, W. J. Romanow, A. E. Sedeño-Cortés, K. Siletti, S. Somasundaram, J. Sulc,  
729 M. Tieu, A. Torkelson, H. Tung, X. Wang, F. Xie, A. M. Yanny, R. Zhang, S. A. Ament, M. M.  
730 Behrens, H. C. Bravo, J. Chun, A. Dobin, J. Gillis, R. Hertzano, P. R. Hof, T. Höllt, G. D. Horwitz,  
731 C. D. Keene, P. V. Kharchenko, A. L. Ko, B. P. Lelieveldt, C. Luo, E. A. Mukamel, A. Pinto-Duarte,  
732 S. Preissl, A. Regev, B. Ren, R. H. Scheuermann, K. Smith, W. J. Spain, O. R. White, C. Koch, M.  
733 Hawrylycz, B. Tasic, E. Z. Macosko, S. A. McCarroll, J. T. Ting, H. Zeng, K. Zhang, G. Feng, J. R.  
734 Ecker, S. Linnarsson, E. S. Lein, Comparative cellular analysis of motor cortex in human, marmoset  
735 and mouse. *Nature*. **598**, 111–119 (2021).

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

- 747 Hinrichs, W. Hlavina, T. Holzer, F. Hsu, A. Hua, T. Hubbard, A. Hunt, I. Jackson, D. B. Jaffe, L. S.  
748 Johnson, M. Jones, T. A. Jones, A. Joy, M. Kamal, E. K. Karlsson, D. Karolchik, A. Kasprzyk, J.  
749 Kawai, E. Keibler, C. Kells, W. J. Kent, A. Kirby, D. L. Kolbe, I. Korf, R. S. Kucherlapati, E. J.  
750 Kulbokas, D. Kulp, T. Landers, J. P. Leger, S. Leonard, I. Letunic, R. Levine, J. Li, M. Li, C. Lloyd,  
751 S. Lucas, B. Ma, D. R. Maglott, E. R. Mardis, L. Matthews, E. Mauceli, J. H. Mayer, M. McCarthy,  
752 W. R. McCombie, S. McLaren, K. McLay, J. D. McPherson, J. Meldrim, B. Meredith, J. P. Mesirov,  
753 W. Miller, T. L. Miner, E. Mongin, K. T. Montgomery, M. Morgan, R. Mott, J. C. Mullikin, D. M.  
754 Muzny, W. E. Nash, J. O. Nelson, M. N. Nhan, R. Nicol, Z. Ning, C. Nusbaum, M. J. O'Connor, Y.  
755 Okazaki, K. Oliver, E. O. Larty, L. Pachter, G. Parra, K. H. Pepin, J. Peterson, P. Pevzner, R. Plumb,  
756 C. S. Pohl, A. Poliakov, T. C. Ponce, C. P. Ponting, S. Potter, M. Quail, A. Reymond, B. A. Roe, K.  
757 M. Roskin, E. M. Rubin, A. G. Rust, R. Santos, V. Sapojnikov, B. Schultz, J. Schultz, M. S.  
758 Schwartz, S. Schwartz, C. Scott, S. Seaman, S. Searle, T. Sharpe, A. Sheridan, R. Shownkeen, S.  
759 Sims, J. B. Singer, G. Slater, A. Smit, D. R. Smith, B. Spencer, A. Stabenau, N. S. Strange-  
760 Thomann, C. Sugnet, M. Suyama, G. Tesler, J. Thompson, D. Torrents, E. Trevaskis, J. Tromp, C.  
761 Ucla, A. U. Vidal, J. P. Vinson, A. C. von Niederhausern, C. M. Wade, M. Wall, R. J. Weber, R. B.  
762 Weiss, M. C. Wendl, A. P. West, K. Wetterstrand, R. Wheeler, S. Whelan, J. Wierzbowski, D.  
763 Willey, S. Williams, R. K. Wilson, E. Winter, K. C. Worley, D. Wyman, S. Yang, S. P. Yang, E. M.  
764 Zdobnov, M. C. Zody, E. S. Lander, C. Mouse Genome Sequencing, Initial sequencing and  
765 comparative analysis of the mouse genome. *Nature*. **420**, 520–562 (2002).
- 766 98. B. Langmead, S. L. Salzberg, Fast gapped-read alignment with Bowtie 2. *Nat. Methods*. **9**, 357–359  
767 (2012).
- 768 99. K. R. Rosenbloom, J. Armstrong, G. P. Barber, J. Casper, H. Clawson, M. Diekhans, T. R. Dreszer,  
769 P. A. Fujita, L. Guruvadoo, M. Haeussler, R. A. Harte, S. Heitner, G. Hickey, A. S. Hinrichs, R.  
770 Hubley, D. Karolchik, K. Learned, B. T. Lee, C. H. Li, K. H. Miga, N. Nguyen, B. Paten, B. J.  
771 Raney, A. F. A. Smit, M. L. Speir, A. S. Zweig, D. Haussler, R. M. Kuhn, W. J. Kent, The UCSC

- 772 Genome Browser database: 2015 update. *Nucleic Acids Res.* **43**, D670–D681 (2015).
- 773 100. H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R.  
774 Durbin, The Sequence Alignment/Map format and SAMtools. *Bioinformatics.* **25**, 2078–2079  
775 (2009).
- 776 101. F. Ramírez, D. P. Ryan, B. Grüning, V. Bhardwaj, F. Kilpert, A. S. Richter, S. Heyne, F. Dündar,  
777 T. Manke, deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic*  
778 *Acids Res.* **44**, W160–5 (2016).
- 779 102. J. M. Granja, M. R. Corces, S. E. Pierce, S. T. Bagdatli, H. Choudhry, H. Y. Chang, W. J.  
780 Greenleaf, ArchR is a scalable software package for integrative single-cell chromatin accessibility  
781 analysis. *Nat. Genet.* **53**, 403–411 (2021).
- 782 103. BRAIN Initiative Cell Census Network (BICCN), A multimodal cell census and atlas of the  
783 mammalian primary motor cortex. *Nature.* **598**, 86–102 (2021).
- 784 104. G. K. Marinov, A. Kundaje, P. Park B. Wold, Large-scale quality analysis of published ChIP-seq  
785 data. *G3 (Bethesda).* **4**, 209–223 (2013).
- 786 105. Q. Li, J. B. Brown, H. Huang, P. J. Bickel, Measuring reproducibility of high-throughput  
787 experiments. *Ann. Appl. Stat.* **5**, 1752–1779 (2011).
- 788 106. A. Frankish, M. Diekhans, A. M. Ferreira, R. Johnson, I. Jungreis, J. Loveland, J. M. Mudge, C.  
789 Sisu, J. Wright, J. Armstrong, I. Barnes, A. Berry, A. Bignell, S. Carbonell Sala, J. Chrast, F.  
790 Cunningham, T. Di Domenico, S. Donaldson, I. T. Fiddes, C. García Girón, J. M. Gonzalez, T.  
791 Grego, M. Hardy, T. Hourlier, T. Hunt, O. G. Izuogu, J. Lagarde, F. J. Martin, L. Martínez, S.  
792 Mohanan, P. Muir, F. C. P. Navarro, A. Parker, B. Pei, F. Pozo, M. Ruffier, B. M. Schmitt, E.  
793 Stapleton, M. M. Suner, I. Sycheva, B. Uszczyńska-Ratajczak, J. Xu, A. Yates, D. Zerbino, Y.  
794 Zhang, B. Aken, J. S. Choudhary, M. Gerstein, R. Guigó, T. J. P. Hubbard, M. Kellis, B. Paten, A.

- 795 Reymond, M. L. Tress, P. Flicek, GENCODE reference annotation for the human and mouse  
796 genomes. *Nucleic Acids Res.* **47**, D766–D773 (2019).
- 797 107. N. A. O’Leary, M. W. Wright, J. R. Brister, S. Ciufu, D. Haddad, R. McVeigh, B. Rajput, B.  
798 Robbertse, B. Smith-White, D. Ako-Adjei, A. Astashyn, A. Badretdin, Y. Bao, O. Blinkova, V.  
799 Brover, V. Chetvernin, J. Choi, E. Cox, O. Ermolaeva, C. M. Farrell, T. Goldfarb, T. Gupta, D. Haft,  
800 E. Hatcher, W. Hlavina, V. S. Joardar, V. K. Kodali, W. Li, D. Maglott, P. Masterson, K. M.  
801 McGarvey, M. R. Murphy, K. O’Neill, S. Pujar, S. H. Rangwala, D. Rausch, L. D. Riddick, C.  
802 Schoch, A. Shkeda, S. S. Storz, H. Sun, F. Thibaud-Nissen, I. Tolstoy, R. E. Tully, A. R. Vatsan, C.  
803 Wallin, D. Webb, W. Wu, M. J. Landrum, A. Kimchi, T. Tatusova, M. DiCuccio, P. Kitts, T. D.  
804 Murphy, K. D. Pruitt, Reference sequence (RefSeq) database at NCBI: current status, taxonomic  
805 expansion, and functional annotation. *Nucleic Acids Res.* **44**, D733–D745 (2016).
- 806 108. W. J. Kent, R. Baertsch, A. Hinrichs, W. Miller, D. Haussler, Evolution’s cauldron: duplication,  
807 deletion, and rearrangement in the mouse and human genomes. *Proc. Natl. Acad. Sci. U. S. A.* **100**,  
808 11484–11489 (2003).
- 809 109. D. Jebb, Z. Huang, M. Pippel, G. M. Hughes, K. Lavrichenko, P. Devanna, S. Winkler, L. S.  
810 Jermin, E. C. Skirmuntt, A. Katzourakis, L. Burkitt-Gray, D. A. Ray, K. A. M. Sullivan, J. G.  
811 Roscito, B. M. Kirilenko, L. M. Dávalos, A. P. Corthals, M. L. Power, G. Jones, R. D. Ransome, D.  
812 K. N. Dechmann, A. G. Locatelli, S. J. Puechmaille, O. Fedrigo, E. D. Jarvis, M. Hiller, S. C.  
813 Vernes, E. W. Myers, E. C. Teeling, Six reference-quality genomes reveal evolution of bat  
814 adaptations. *Nature.* **583**, 578–584 (2020).
- 815 110. Y. Zhang, T. Liu, C. A. Meyer, J. Eeckhoute, D. S. Johnson, B. E. Bernstein, C. Nussbaum, R. M.  
816 Myers, M. Brown, W. Li, X. S. Liu, Model-based Analysis of ChIP-Seq (MACS). *Genome Biol.* **9**,  
817 R137 (2008).

- 818 111. M. Ghandi, M. Mohammad-Noori, N. Ghareghani, D. Lee, L. Garraway, M. A. Beer, gkmSVM:  
819 an R package for gapped-kmer SVM. *Bioinformatics*. **32**, 2205–2207 (2016).
- 820 112. W. Huber, V. J. Carey, R. Gentleman, S. Anders, M. Carlson, B. S. Carvalho, H. C. Bravo, S.  
821 Davis, L. Gatto, T. Girke, R. Gottardo, F. Hahne, K. D. Hansen, R. A. Irizarry, M. Lawrence, M. I.  
822 Love, J. MacDonald, V. Obenchain, A. K. Oleś, H. Pagès, A. Reyes, P. Shannon, G. K. Smyth, D.  
823 Tenenbaum, L. Waldron, M. Morgan, Orchestrating high-throughput genomic analysis with  
824 Bioconductor. *Nat. Methods*. **12**, 115–121 (2015).
- 825 113. A. Khan, R. Riudavets Puig, P. Boddie, A. Mathelier, BiasAway: command-line and web server  
826 to generate nucleotide composition-matched DNA background sequences. *Bioinformatics*. **37**, 1607–  
827 1609 (2021).
- 828 114. R. Worsley Hunt, A. Mathelier, L. Del Peso, W. W. Wasserman, Improving analysis of  
829 transcription factor binding sites within ChIP-Seq data based on topological motif enrichment. *BMC*  
830 *Genom.* **15**, 472 (2014).
- 831 115. B. Alipanahi, A. DeLong, M. T. Weirauch, B. J. Frey, Predicting the sequence specificities of  
832 DNA- and RNA-binding proteins by deep learning. *Nat. Biotechnol.* **33**, 831–838 (2015).
- 833 116. F. Chollet, Keras (2015). <https://keras.io>.
- 834 117. J. Bergstra, O. Breuleux, F. Bastien, P. Lamblin, R. Pascanu, G. Desjardins, J. Turian, D. Warde-  
835 Farley, Y. Bengio, Theano: A CPU and GPU Math Compiler in Python. *Proceedings of the Python*  
836 *for scientific computing conference (SciPy)*. **4**, 1–7 (2010).
- 837 118. M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J.  
838 Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, R. Jozefowicz, Y. Jia,  
839 L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, M. Schuster, R. Monga, S. Moore, D. Murray, C.  
840 Olah, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F.

- 841 Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, X. Zheng, TensorFlow: a system  
842 for large-scale machine learning. *Proceedings of the 12th USENIX conference on Operating Systems  
843 Design and Implementation*. **12**, 262-283 (2016).
- 844 119. F. Pedregosa, G. Varoquaux, Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **12**,  
845 2825–2830 (2011).
- 846 120. J. Grau, I. Grosse, J. Keilwagen, PRROC: Computing and visualizing Precision-recall and  
847 receiver operating characteristic curves in R. *Bioinformatics*. **31**, 2595–2597 (2015).
- 848 121. K. He, X. Zhang, S. Ren, J. Sun, Delving Deep into Rectifiers: Surpassing Human-Level  
849 Performance on ImageNet Classification. *Proceedings of the IEEE International Conference on  
850 Computer Vision and Pattern Recognition*. 1026–1034 (2015).
- 851 122. X. Glorot, Y. Bengio, Understanding the difficulty of training deep feedforward neural networks.  
852 *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics,  
853 Proceedings of Machine Learning Research*. **9**, 249–256 (2010).
- 854 123. L. N. Smith, Cyclical Learning Rates for Training Neural Networks. *2017 IEEE Winter  
855 Conference on Applications of Computer Vision (WACV)*. 464–472 (2017).
- 856 124. A. Shrikumar, K. Tian, A. Shcherbina, Ž. Avsec, A. Banerjee, M. Sharmin, S. Nair, A. Kundaje,  
857 <https://arxiv.org/abs/1811.00416>.
- 858 125. S. Gupta, J. A. Stamatoyannopoulos, T. L. Bailey, W. Noble, Quantifying similarity between  
859 motifs. *Genome Biol.* **8**, R24 (2007).
- 860 126. I. V. Kulakovskiy, I. E. Vorontsov, I. S. Yevshin, R. N. Sharipov, A. D. Fedorova, E. I.  
861 Rumynskiy, Y. A. Medvedeva, A. Magana-Mora, V. B. Bajic, D. A. Papatsenko, F. A. Kolpakov, V.  
862 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 *Consortium*.
- 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 Consortium*.
- 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).

- 930 150. S. Shim, K. Y. Kwan, M. Li, V. Lefebvre, N. Sestan, Cis-regulatory control of corticospinal  
931 system development and evolution. *Nature*. **486**, 74–79 (2012).
- 932 151. T. Hisaoka, Y. Nakamura, E. Senba, Y. Morikawa, The forkhead transcription factors, Foxp1 and  
933 Foxp2, identify different subpopulations of projection neurons in the mouse cerebral cortex.  
934 *Neuroscience*. **166**, 551–563 (2010).
- 935 152. K. Ohi, R. Hashimoto, Y. Yasuda, M. Kiribayashi, N. Iike, T. Yoshida, M. Azechi, K. Ikezawa,  
936 H. Takahashi, T. Morihara, R. Ishii, S. Tagami, M. Iwase, M. Okochi, K. Kamino, H. Kazui, T.  
937 Tanaka, T. Kudo, M. Takeda, TATA box-binding protein gene is associated with risk for  
938 schizophrenia, age at onset and prefrontal function. *Genes Brain Behav.* **8**, 473–480 (2009).
- 939 153. B. Muralidharan, Z. Khatri, U. Maheshwari, R. Gupta, B. Roy, S. J. Pradhan, K. Karmodiya, H.  
940 Padmanabhan, A. S. Shetty, C. Balaji, U. Kolthur-Seetharam, J. D. Macklis, S. Galande, S. Tole,  
941 LHX2 Interacts with the NuRD Complex and Regulates Cortical Neuron Subtype Determinants  
942 Fezf2 and Sox11. *J. Neurosci.* **37**, 194–203 (2017).
- 943 154. H. Yang, J. Kim, Y. Kim, S.-W. Jang, N. Sestan, S. Shim, Cux2 expression regulated by Lhx2 in  
944 the upper layer neurons of the developing cortex. *Biochem. Biophys. Res. Commun.* **521**, 874–879  
945 (2020).
- 946 155. The Human Protein Atlas. <http://www.proteinatlas.org>.
- 947 156. Y. Xue, C. Guo, F. Hu, W. Zhu, S. Mao, PPARA/RXRA signalling regulates the fate of hepatic  
948 non-esterified fatty acids in a sheep model of maternal undernutrition. *Biochim. Biophys. Acta Mol.*  
949 *Cell Biol. Lipids.* **1865**, 158548 (2020).
- 950 157. F. Stossi, R. D. Dandekar, H. Johnson, P. Lavere, C. E. Foulds, M. G. Mancini, M. A. Mancini,  
951 Tributyltin chloride (TBT) induces RXRA down-regulation and lipid accumulation in human liver  
952 cells. *PLoS One.* **14**, e0224405 (2019).

- 953 158. K. J. Wangensteen, S. Zhang, L. E. Greenbaum, K. H. Kaestner, A genetic screen reveals Foxa3  
954 and TNFR1 as key regulators of liver repopulation. *Genes Dev.* **29**, 904–909 (2015).
- 955 159. J. R. Friedman, K. H. Kaestner, The Foxa family of transcription factors in development and  
956 metabolism. *Cell. Mol. Life Sci.* **63**, 2317–2328 (2006).
- 957 160. H. Schrem, J. Klempnauer, J. Borlak, Liver-enriched transcription factors in liver function and  
958 development. Part II: the C/EBPs and D site-binding protein in cell cycle control, carcinogenesis,  
959 circadian gene regulation, liver regeneration, apoptosis, and liver-specific gene regulation.  
960 *Pharmacol. Rev.* **56**, 291–330 (2004).
- 961 161. P. Hatzis, I. Kyrmizi, I. Talianidis, Mitogen-activated protein kinase-mediated disruption of  
962 enhancer-promoter communication inhibits hepatocyte nuclear factor 4alpha expression. *Mol. Cell.*  
963 *Biol.* **26**, 7017–7029 (2006).
- 964 162. J. B. Nielsen, O. Rom, I. Surakka, S. E. Graham, W. Zhou, T. Roychowdhury, L. G. Fritsche, S.  
965 A. Gagliano Taliun, C. Sidore, Y. Liu, M. E. Gabrielsen, A. H. Skogholt, B. Wolford, W. Overton,  
966 Y. Zhao, J. Chen, H. Zhang, W. E. Hornsby, A. Acheampong, A. Grooms, A. Schaefer, G. J. M.  
967 Zajac, L. Villacorta, J. Zhang, B. Brumpton, M. Løset, V. Rai, P. R. Lundegaard, M. S. Olesen, K.  
968 D. Taylor, N. D. Palmer, Y.-D. Chen, S. H. Choi, S. A. Lubitz, P. T. Ellinor, K. C. Barnes, M. Daya,  
969 N. Rafaels, S. T. Weiss, J. Lasky-Su, R. P. Tracy, R. S. Vasani, L. A. Cupples, R. A. Mathias, L. R.  
970 Yanek, L. C. Becker, P. A. Peyser, L. F. Bielak, J. A. Smith, S. Aslibekyan, B. A. Hidalgo, D. K.  
971 Arnett, M. R. Irvin, J. G. Wilson, S. K. Musani, A. Correa, S. S. Rich, X. Guo, J. I. Rotter, B. A.  
972 Konkle, J. M. Johnsen, A. E. Ashley-Koch, M. J. Telen, V. A. Sheehan, J. Blangero, J. E. Curran, J.  
973 M. Peralta, C. Montgomery, W. H.-H. Sheu, R.-H. Chung, K. Schwander, S. M. Nouraei, V. R.  
974 Gordeuk, Y. Zhang, C. Kooperberg, A. P. Reiner, R. D. Jackson, E. R. Bleecker, D. A. Meyers, X.  
975 Li, S. Das, K. Yu, J. LeFaive, A. Smith, T. Blackwell, D. Taliun, S. Zollner, L. Forer, S. Schoenherr,  
976 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. Fougousse, 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

- 999 vasculature development. *Int. J. Ophthalmol.* **13**, 231–238 (2020).
- 1000 171. R. Ratnapriya, O. A. Sosina, M. R. Starostik, M. Kwicklis, R. J. Kapphahn, L. G. Fritsche, A.  
1001 Walton, M. Arvanitis, L. Gieser, A. Pietraszkiewicz, S. R. Montezuma, E. Y. Chew, A. Battle, G. R.  
1002 Abecasis, D. A. Ferrington, N. Chatterjee, A. Swaroop, Retinal transcriptome and eQTL analyses  
1003 identify genes associated with age-related macular degeneration. *Nat. Genet.* **51**, 606–610 (2019).
- 1004 172. A.-G. Wang, C.-H. Chen, C.-W. Yang, M.-Y. Yen, W.-M. Hsu, J.-H. Liu, M.-J. Fann, Change of  
1005 gene expression profiles in the retina following optic nerve injury. *Brain Res. Mol. Brain Res.* **101**,  
1006 82–92 (2002).
- 1007 173. A. Swaroop, D. Kim, D. Forrest, Transcriptional regulation of photoreceptor development and  
1008 homeostasis in the mammalian retina. *Nat. Rev. Neurosci.* **11**, 563–576 (2010).
- 1009 174. P.-H. Fabre, L. Hautier, D. Dimitrov, E. J. P. Douzery, A glimpse on the pattern of rodent  
1010 diversification: a phylogenetic approach. *BMC Evol. Biol.* **12**, 88 (2012).
- 1011 175. T. C. Giarla, J. A. Esselstyn, The Challenges of Resolving a Rapid, Recent Radiation: Empirical  
1012 and Simulated Phylogenomics of Philippine Shrews. *Syst. Biol.* **64**, 727–740 (2015).
- 1013 176. G. Shi, Z. Zhang, Rap2B promotes the proliferation and migration of human glioma cells via  
1014 activation of the ERK pathway. *Oncol. Lett.* **21**, 314 (2021).
- 1015 177. S. P. F. Ensign, A. Roos, I. T. Mathews, H. D. Dhruv, S. Tuncali, J. N. Sarkaria, M. H. Symons,  
1016 J. C. Loftus, M. E. Berens, N. L. Tran, SGEF Is Regulated via TWEAK/Fn14/NF- $\kappa$ B Signaling and  
1017 Promotes Survival by Modulation of the DNA Repair Response to Temozolomide. *Mol. Cancer Res.*  
1018 **14**, 302–312 (2016).
- 1019 178. H. Liu, Y.-N. Lu, T. Paul, G. Periz, M. T. Banco, A. R. Ferré-D’Amaré, J. D. Rothstein, L. R.  
1020 Hayes, S. Myong, J. Wang, A Helicase Unwinds Hexanucleotide Repeat RNA G-Quadruplexes and

- 1021 Facilitates Repeat-Associated Non-AUG Translation. *J. Am. Chem. Soc.* **143**, 7368–7379 (2021).
- 1022 179. Y. Shangguan, X. Xu, B. Ganbat, Y. Li, W. Wang, Y. Yang, X. Lu, C. Du, X. Tian, X. Wang,  
1023 CNTNAP4 Impacts Epilepsy Through GABAA Receptors Regulation: Evidence From Temporal  
1024 Lobe Epilepsy Patients and Mouse Models. *Cereb. Cortex.* **28**, 3491–3504 (2018).
- 1025 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,  
1026 S. Oiso, K. Nishida, S. Kuchiiwa, K. Watanabe, K. Yamada, Y. Takeda, Z.-C. Xiao, Q.-H. Ma,  
1027 Caspr4 interaction with LNX2 modulates the proliferation and neuronal differentiation of mouse  
1028 neural progenitor cells. *Stem Cells Dev.* **24**, 640–652 (2015).
- 1029 181. C. Sun, J. Song, Y. Jiang, C. Zhao, J. Lu, Y. Li, Y. Wang, M. Gao, J. Xi, S. Luo, M. Li, K.  
1030 Donaldson, S. N. Oprescu, T. P. Slavin, S. Lee, P. L. Magoulas, A. M. Lewis, L. Emrick, S. R.  
1031 Lalani, Z. Niu, M. L. Landsverk, M. Walkiewicz, R. E. Person, H. Mei, J. A. Rosenfeld, Y. Yang, A.  
1032 Antonellis, Y.-M. Hou, J. Lin, V. W. Zhang, Loss-of-function mutations in Lysyl-tRNA synthetase  
1033 cause various leukoencephalopathy phenotypes. *Neurol Genet.* **5**, e565 (2019).
- 1034 182. H. J. McMillan, P. Humphreys, A. Smith, J. Schwartzenruber, P. Chakraborty, D. E. Bulman, C.  
1035 L. Beaulieu, FORGE Canada Consortium, J. Majewski, K. M. Boycott, M. T. Geraghty, Congenital  
1036 Visual Impairment and Progressive Microcephaly Due to Lysyl-Transfer Ribonucleic Acid (RNA)  
1037 Synthetase (KARS) Mutations: The Expanding Phenotype of Aminoacyl-Transfer RNA Synthetase  
1038 Mutations in Human Disease. *J. Child Neurol.* **30**, 1037–1043 (2015).
- 1039 183. C. C. Hong, A. T. Tang, M. R. Detter, J. P. Choi, R. Wang, X. Yang, A. A. Guerrero, C. F.  
1040 Wittig, N. Hobson, R. Girard, R. Lightle, T. Moore, R. Shenkar, S. P. Polster, L. M. Goddard, A. A.  
1041 Ren, N. A. Leu, S. Sterling, J. Yang, L. Li, M. Chen, P. Mericko-Ishizuka, L. E. Dow, H. Watanabe,  
1042 M. Schwaninger, W. Min, D. A. Marchuk, X. Zheng, I. A. Awad, M. L. Kahn, Cerebral cavernous  
1043 malformations are driven by ADAMTS5 proteolysis of versican. *J. Exp. Med.* **217**, e20200140



- 1044 (2020).
- 1045 184. H. Kumon, Y. Yoshino, Y. Funahashi, H. Mori, M. Ueno, Y. Ozaki, K. Yamazaki, S. Ochi, T.  
1046 Mori, J.-I. Iga, M. Nagai, M. Nomoto, S.-I. Ueno, PICALM mRNA Expression in the Blood of  
1047 Patients with Neurodegenerative Diseases and Geriatric Depression. *J. Alzheimers. Dis.* **79**, 1055–  
1048 1062 (2021).
- 1049 185. E. Cooney, W. Bi, A. E. Schlesinger, S. Vinson, L. Potocki, Novel EED mutation in patient with  
1050 Weaver syndrome. *Am. J. Med. Genet. A.* **173**, 541–545 (2017).
- 1051 186. J. Wang, L. Yang, C. Dong, J. Wang, L. Xu, Y. Qiu, Q. Weng, C. Zhao, M. Xin, Q. R. Lu, EED-  
1052 mediated histone methylation is critical for CNS myelination and remyelination by inhibiting WNT,  
1053 BMP, and senescence pathways. *Sci Adv.* **6**, eaaz6477 (2020).
- 1054 187. G. Helman, A. Zerem, A. Almad, J. L. Hacker, S. Woidill, S. Sase, A. N. LeFevre, J. Ekstein, M.  
1055 M. Johansson, C. A. Stutterd, R. J. Taft, C. Simons, J. B. Grinspan, A. Pizzino, J. L. Schmidt, B.  
1056 Harding, Y. Hirsch, A. N. Viaene, A. Fattal-Valevski, A. Vanderver, Further Delineation of the  
1057 Clinical and Pathologic Features of HIKESHI-Related Hypomyelinating Leukodystrophy. *Pediatr.*  
1058 *Neurol.* **121**, 11–19 (2021).
- 1059 188. S. Edvardson, S. Kose, C. Jalas, A. Fattal-Valevski, A. Watanabe, Y. Ogawa, H. Mamada, A. M.  
1060 Fedick, S. Ben-Shachar, N. R. Treff, A. Shaag, S. Bale, J. Gärtner, N. Imamoto, O. Elpeleg,  
1061 Leukoencephalopathy and early death associated with an Ashkenazi-Jewish founder mutation in the  
1062 Hikeshi gene. *J. Med. Genet.* **53**, 132–137 (2016).
- 1063 189. A. Oguro-Ando, R. A. Bamford, W. Sital, J. J. Sprengers, A. Zuko, J. M. Matser, H. Oppelaar, A.  
1064 Sarabdjitsingh, M. Joëls, J. P. H. Burbach, M. J. Kas, Cntn4, a risk gene for neuropsychiatric  
1065 disorders, modulates hippocampal synaptic plasticity and behavior. *Transl. Psychiatry.* **11**, 106  
1066 (2021).

- 1067 190. J. Hu, J. Liao, M. Sathanoori, S. Kochmar, J. Sebastian, S. A. Yatsenko, U. Surti, CNTN6 copy  
1068 number variations in 14 patients: a possible candidate gene for neurodevelopmental and  
1069 neuropsychiatric disorders. *J. Neurodev. Disord.* **7**, 26 (2015).
- 1070 191. L. M. Hansford, S. A. Smith, M. Haber, M. D. Norris, B. Cheung, G. M. Marshall, Cloning and  
1071 characterization of the human neural cell adhesion molecule, CNTN4 (alias BIG-2). *Cytogenet.*  
1072 *Genome Res.* **101**, 17–23 (2003).
- 1073 192. Y. Kamei, O. Tsutsumi, Y. Taketani, K. Watanabe, cDNA cloning and chromosomal localization  
1074 of neural adhesion molecule NB-3 in human. *J. Neurosci. Res.* **51**, 275–283 (1998).
- 1075 193. K. Baker, S. L. Gordon, H. Melland, F. Bumbak, D. J. Scott, T. J. Jiang, D. Owen, B. J. Turner,  
1076 S. G. Boyd, M. Rossi, M. Al-Raqad, O. Elpeleg, D. Peck, G. M. S. Mancini, M. Wilke, M. Zollino,  
1077 G. Marangi, H. Weigand, I. Borggraefe, T. Haack, Z. Stark, S. Sadedin, Broad Center for Mendelian  
1078 Genomics, T. Y. Tan, Y. Jiang, R. A. Gibbs, S. Ellingwood, M. Amaral, W. Kelley, M. A. Kurian,  
1079 M. A. Cousin, F. L. Raymond, SYT1-associated neurodevelopmental disorder: a case series. *Brain.*  
1080 **141**, 2576–2591 (2018).
- 1081 194. A. Shukla, R. P. Saneto, M. Hebbar, G. Mirzaa, K. M. Girisha, A neurodegenerative  
1082 mitochondrial disease phenotype due to biallelic loss-of-function variants in PNPLA8 encoding  
1083 calcium-independent phospholipase A2 $\gamma$ . *Am. J. Med. Genet. A.* **176**, 1232–1237 (2018).
- 1084 195. M. Paucar, R. Ågren, T. Li, S. Lissmats, Å. Bergendal, J. Weinberg, D. Nilsson, I. Savichetva, K.  
1085 Sahlholm, J. Nilsson, P. Svenningsson, V374A KCND3 Pathogenic Variant Associated With  
1086 Paroxysmal Ataxia Exacerbations. *Neurol Genet.* **7**, e546 (2021).
- 1087 196. G. Skariah, K. J. Perry, J. Drnevich, J. J. Henry, S. Ceman, RNA helicase Mov10 is essential for  
1088 gastrulation and central nervous system development. *Dev. Dyn.* **247**, 660–671 (2018).
- 1089 197. Q. Xiao, M. Dong, F. Cheng, F. Mao, W. Zong, K. Wu, H. Wang, R. Xie, B. Wang, T. Lei, D.

- 1090 Guo, LRIG2 promotes the proliferation and cell cycle progression of glioblastoma cells in vitro and  
1091 in vivo through enhancing PDGFR $\beta$  signaling. *Int. J. Oncol.* **53**, 1069–1082 (2018).
- 1092 198. A.-R. Lee, K. W. Ko, H. Lee, Y.-S. Yoon, M.-R. Song, C.-S. Park, Putative Cell Adhesion  
1093 Membrane Protein Vstm5 Regulates Neuronal Morphology and Migration in the Central Nervous  
1094 System. *J. Neurosci.* **36**, 10181–10197 (2016).
- 1095 199. A. Fattal-Valevski, L. Ben Sira, T. Lerman-Sagie, R. Strausberg, A. Bloch-Mimouni, S.  
1096 Edvardson, R. Kaufman, V. Chernuha, N. Schneebaum Sender, G. Heimer, B. Ben Zeev, Delineation  
1097 of the phenotype of MED17-related disease in Caucasus-Jewish families. *Eur. J. Paediatr. Neurol.*  
1098 **32**, 40–45 (2021).
- 1099 200. I. R. Raslan, P. C. A. de Assis Pereira Matos, V. Boaratti Ciarlariello, K. H. Daghasanli, A. B. R.  
1100 Rosa, J. H. Arita, C. S. Aranda, O. G. P. Barsottini, J. L. Pedroso, Beyond Typical Ataxia  
1101 Telangiectasia: How to Identify the Ataxia Telangiectasia-Like Disorders. *Mov Disord Clin Pract.* **8**,  
1102 118–125 (2021).
- 1103 201. C. S. Subhramanyam, Q. Cao, C. Wang, Z. S. L. Heng, Z. Zhou, Q. Hu, Role of PIWI-like 4 in  
1104 modulating neuronal differentiation from human embryonal carcinoma cells. *RNA Biol.* **17**, 1613–  
1105 1624 (2020).
- 1106 202. G. S. Kinker, L. H. Ostrowski, P. A. C. Ribeiro, R. Chanoch, S. M. Muxel, I. Tirosh, G. Spadoni,  
1107 S. Rivara, V. R. Martins, T. G. Santos, R. P. Markus, P. A. C. M. Fernandes, MT1 and MT2  
1108 melatonin receptors play opposite roles in brain cancer progression. *J. Mol. Med.* **99**, 289–301  
1109 (2021).
- 1110 203. M.-Y. Ou, X.-C. Ju, Y.-J. Cai, X.-Y. Sun, J.-F. Wang, X.-Q. Fu, Q. Sun, Z.-G. Luo,  
1111 Heterogeneous nuclear ribonucleoprotein A3 controls mitotic progression of neural progenitors via  
1112 interaction with cohesin. *Development.* **147**, dev185132 (2020).

- 1113 204. M. Zech, F. Castrop, B. Schormair, A. Jochim, T. Wieland, N. Gross, P. Lichtner, A. Peters, C.  
1114 Gieger, T. Meitinger, T. M. Strom, K. Oexle, B. Haslinger, J. Winkelmann, DYT16 revisited: exome  
1115 sequencing identifies PRKRA mutations in a European dystonia family. *Mov. Disord.* **29**, 1504–  
1116 1510 (2014).
- 1117 205. J. Berciano, K. Peeters, A. García, T. López-Alburquerque, E. Gallardo, A. Hernández-Fabián, A.  
1118 L. Pelayo-Negro, E. De Vriendt, J. Infante, A. Jordanova, NEFL N98S mutation: another cause of  
1119 dominant intermediate Charcot-Marie-Tooth disease with heterogeneous early-onset phenotype. *J.*  
1120 *Neurol.* **263**, 361–369 (2016).
- 1121 206. J. Berciano, A. García, E. Gallardo, K. Peeters, A. L. Pelayo-Negro, S. Álvarez-Paradelo, J.  
1122 Gazulla, M. Martínez-Tames, J. Infante, A. Jordanova, Intermediate Charcot-Marie-Tooth disease:  
1123 an electrophysiological reappraisal and systematic review. *J. Neurol.* **264**, 1655–1677 (2017).
- 1124 207. A. Jimenez-Pascual, J. S. Hale, A. Kordowski, J. Pugh, D. J. Silver, D. Bayik, G. Roversi, T. J.  
1125 Alban, S. Rao, R. Chen, T. M. McIntyre, G. Colombo, G. Taraboletti, K. O. Holmberg, K. Forsberg-  
1126 Nilsson, J. D. Lathia, F. A. Siebzehnruhl, ADAMDEC1 Maintains a Growth Factor Signaling Loop  
1127 in Cancer Stem Cells. *Cancer Discov.* **9**, 1574–1589 (2019).
- 1128 208. A. Badaloni, F. Casoni, L. Croci, F. Chiara, A. Bizzoca, G. Gennarini, O. Cremona, R. Hawkes,  
1129 G. G. Consalez, Dynamic Expression and New Functions of Early B Cell Factor 2 in Cerebellar  
1130 Development. *Cerebellum.* **18**, 999–1010 (2019).
- 1131 209. M. Wagner, Y. Skorobogatko, B. Pode-Shakked, C. M. Powell, B. Alhaddad, A. Seibt, O. Barel,  
1132 G. Heimer, C. Hoffmann, L. A. Demmer, Y. Perilla-Young, M. Remke, D. Wieczorek, T.  
1133 Navaratnarajah, P. Lichtner, D. Klee, H. E. Shamseldin, F. Al Mutairi, E. Mayatepek, T. Strom, T.  
1134 Meitinger, F. S. Alkuraya, Y. Anikster, A. R. Saltiel, F. Distelmaier, Bi-allelic Variants in  
1135 RALGAPA1 Cause Profound Neurodevelopmental Disability, Muscular Hypotonia, Infantile

- 1136 Spasms, and Feeding Abnormalities. *Am. J. Hum. Genet.* **106**, 246–255 (2020).
- 1137 210. T. Guran, G. Yesil, S. Turan, Z. Atay, E. Bozkurtlar, A. Aghayev, S. Gul, I. Tinay, B. Aru, S.  
1138 Arslan, M. K. Koroglu, F. Ercan, G. Y. Demirel, F. S. Eren, B. Karademir, A. Bereket, PPP2R3C  
1139 gene variants cause syndromic 46,XY gonadal dysgenesis and impaired spermatogenesis in humans.  
1140 *Eur. J. Endocrinol.* **180**, 291–309 (2019).
- 1141 211. A. Carré, G. Szinnai, M. Castanet, S. Sura-Trueba, E. Tron, I. Broutin-L’Hermite, P. Barat, C.  
1142 Goizet, D. Lacombe, M.-L. Moutard, C. Raybaud, C. Raynaud-Ravni, S. Romana, H. Ythier, J.  
1143 Léger, M. Polak, Five new TTF1/NKX2.1 mutations in brain-lung-thyroid syndrome: rescue by  
1144 PAX8 synergism in one case. *Hum. Mol. Genet.* **18**, 2266–2276 (2009).
- 1145 212. W. Jarrar, J. M. Dias, J. Ericson, H.-H. Arnold, A. Holz, Nkx2.2 and Nkx2.9 are the key  
1146 regulators to determine cell fate of branchial and visceral motor neurons in caudal hindbrain. *PLoS*  
1147 *One.* **10**, e0124408 (2015).
- 1148 213. V. Weitensteiner, R. Zhang, J. Bungenberg, M. Marks, J. Gehlen, D. J. Ralser, A. C. Hilger, A.  
1149 Sharma, J. Schumacher, U. Gembruch, W. M. Merz, A. Becker, J. Altmüller, H. Thiele, B. G.  
1150 Herrmann, B. Odermatt, M. Ludwig, H. Reutter, Exome sequencing in syndromic brain  
1151 malformations identifies novel mutations in ACTB, and SLC9A6, and suggests BAZ1A as a new  
1152 candidate gene. *Birth Defects Res.* **110**, 587–597 (2018).
- 1153 214. E. Ponzi, M. Gentile, E. Agolini, E. Matera, R. Palumbi, A. L. Buonadonna, A. Peschechera, A.  
1154 Gabellone, M. F. Antonucci, L. Margari, 14q12q13.2 microdeletion syndrome: Clinical  
1155 characterization of a new patient, review of the literature, and further evidence of a candidate region  
1156 for CNS anomalies. *Mol Genet Genomic Med.* **8**, e1289 (2020).
- 1157 215. T. Suzuki, T. Suzuki, M. Raveau, N. Miyake, G. Sudo, Y. Tsurusaki, T. Watanabe, Y. Sugaya, T.  
1158 Tatsukawa, E. Mazaki, A. Shimohata, I. Kushima, B. Aleksic, T. Shiino, T. Toyota, Y. Iwayama, K.

- 1159 Nakaoka, I. Ohmori, A. Sasaki, K. Watanabe, S. Hirose, S. Kaneko, Y. Inoue, T. Yoshikawa, N.  
1160 Ozaki, M. Kano, T. Shimoji, N. Matsumoto, K. Yamakawa, A recurrent PJA1 variant in  
1161 trigonocephaly and neurodevelopmental disorders. *Ann Clin Transl Neurol.* **7**, 1117–1131 (2020).
- 1162 216. S. R. F. Twigg, R. Kan, C. Babbs, E. G. Bochukova, S. P. Robertson, S. A. Wall, G. M. Morriss-  
1163 Kay, A. O. M. Wilkie, Mutations of ephrin-B1 (EFNB1), a marker of tissue boundary formation,  
1164 cause craniofrontonasal syndrome. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 8652–8657 (2004).
- 1165 217. J. Y. Han, H. J. Kim, J. H. Jang, I. G. Lee, J. Park, Trio-Based Whole-Exome Sequencing  
1166 Identifies a De novo EFNB1 Mutation as a Genetic Cause in Female Infant With Brain Anomaly and  
1167 Developmental Delay. *Front Pediatr.* **8**, 461 (2020).
- 1168 218. G. Zanni, Y. Saillour, M. Nagara, P. Billuart, L. Castelnaud, C. Moraine, L. Faivre, E. Bertini, A.  
1169 Durr, A. Guichet, D. Rodriguez, V. des Portes, C. Beldjord, J. Chelly, Oligophrenin 1 mutations  
1170 frequently cause X-linked mental retardation with cerebellar hypoplasia. *Neurology.* **65**, 1364–1369  
1171 (2005).
- 1172 219. A. Hoischen, B. W. M. van Bon, B. Rodríguez-Santiago, C. Gilissen, L. E. L. M. Vissers, P. de  
1173 Vries, I. Janssen, B. van Lier, R. Hastings, S. F. Smithson, R. Newbury-Ecob, S. Kjaergaard, J.  
1174 Goodship, R. McGowan, D. Bartholdi, A. Rauch, M. Peippo, J. M. Cobben, D. Wieczorek, G.  
1175 Gillessen-Kaesbach, J. A. Veltman, H. G. Brunner, B. B. B. A. de Vries, De novo nonsense  
1176 mutations in ASXL1 cause Bohring-Opitz syndrome. *Nat. Genet.* **43**, 729–731 (2011).
- 1177 220. A. H. Alsabban, M. Morikawa, Y. Tanaka, Y. Takei, N. Hirokawa, Kinesin Kif3b mutation  
1178 reduces NMDAR subunit NR2A trafficking and causes schizophrenia-like phenotypes in mice.  
1179 *EMBO J.* **39**, e101090 (2020).
- 1180 221. L. Adnani, R. Dixit, X. Chen, A. Balakrishnan, H. Modi, Y. Touahri, C. Logan, C. Schuurmans,  
1181 Plag1 and Plag2 have overlapping and distinct functions in telencephalic development. *Biol. Open.*

- 1182 7, bio038661 (2018).
- 1183 222. J. Wu, G. Lu, Multiple functions of TBCK protein in neurodevelopment disorders and tumors.  
1184 *Oncol. Lett.* **21**, 17 (2021).
- 1185 223. E. J. Bhoj, D. Li, M. Harr, S. Edvardson, O. Elpeleg, E. Chisholm, J. Juusola, G. Douglas, M. J.  
1186 Guillen Sacoto, K. Siquier-Pernet, A. Saadi, C. Bole-Feysot, P. Nitschke, A. Narravula, M. Walke,  
1187 M. B. Horner, D.-L. Day-Salvatore, P. Jayakar, S. A. S. Vergano, M. A. Tarnopolsky, M. Hegde, L.  
1188 Colleaux, P. Crino, H. Hakonarson, Mutations in TBCK, Encoding TBC1-Domain-Containing  
1189 Kinase, Lead to a Recognizable Syndrome of Intellectual Disability and Hypotonia. *Am. J. Hum.*  
1190 *Genet.* **98**, 782–788 (2016).
- 1191 224. T. Avşar, Ş. Çaliş, B. Yilmaz, G. Demirci Otluoğlu, C. Holyavkin, T. Kiliç, Genome-wide  
1192 identification of Chiari malformation type I associated candidate genes and chromosomal variations.  
1193 *Turk. J. Biol.* **44**, 449–456 (2020).
- 1194 225. H. Behesti, T. R. Fore, P. Wu, Z. Horn, M. Leppert, C. Hull, M. E. Hatten, ASTN2 modulates  
1195 synaptic strength by trafficking and degradation of surface proteins. *Proc. Natl. Acad. Sci. U. S. A.*  
1196 **115**, E9717–E9726 (2018).
- 1197 226. A. P. Chiang, J. S. Beck, H.-J. Yen, M. K. Tayeh, T. E. Scheetz, R. E. Swiderski, D. Y.  
1198 Nishimura, T. A. Braun, K.-Y. A. Kim, J. Huang, K. Elbedour, R. Carmi, D. C. Slusarski, T. L.  
1199 Casavant, E. M. Stone, V. C. Sheffield, Homozygosity mapping with SNP arrays identifies TRIM32,  
1200 an E3 ubiquitin ligase, as a Bardet-Biedl syndrome gene (BBS11). *Proc. Natl. Acad. Sci. U. S. A.*  
1201 **103**, 6287–6292 (2006).
- 1202 227. V. Saccone, M. Palmieri, L. Passamano, G. Piluso, G. Meroni, L. Politano, V. Nigro, Mutations  
1203 that impair interaction properties of TRIM32 associated with limb-girdle muscular dystrophy 2H.  
1204 *Hum. Mutat.* **29**, 240–247 (2008).

- 1205 228. C. J. Schoen, S. B. Emery, M. C. Thorne, H. R. Ammana, E. Sliwerska, J. Arnett, M. Hortsch, F.  
1206 Hannan, M. Burmeister, M. M. Lesperance, Increased activity of Diaphanous homolog 3  
1207 (DIAPH3)/diaphanous causes hearing defects in humans with auditory neuropathy and in  
1208 *Drosophila*. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 13396–13401 (2010).
- 1209 229. E. O.-C. Lau, D. Damiani, G. Chehade, N. Ruiz-Reig, R. Saade, Y. Jossin, M. Aittaleb, O.  
1210 Schakman, N. Tajeddine, P. Gailly, F. Tissir, DIAPH3 deficiency links microtubules to mitotic  
1211 errors, defective neurogenesis, and brain dysfunction. *Elife*. **10**, e61974 (2021).
- 1212 230. M. E. C. Meuwissen, R. Schot, S. Buta, G. Oudesluijs, S. Tinschert, S. D. Speer, Z. Li, L. van  
1213 Unen, D. Heijman, T. Goldmann, M. H. Lequin, J. M. Kros, W. Stam, M. Hermann, R. Willemsen,  
1214 R. W. W. Brouwer, W. F. J. Van IJcken, M. Martin-Fernandez, I. de Coo, J. Dudink, F. A. T. de  
1215 Vries, A. Bertoli Avella, M. Prinz, Y. J. Crow, F. W. Verheijen, S. Pellegrini, D. Bogunovic, G. M.  
1216 S. Mancini, Human USP18 deficiency underlies type 1 interferonopathy leading to severe pseudo-  
1217 TORCH syndrome. *J. Exp. Med.* **213**, 1163–1174 (2016).
- 1218 231. A. T. Vulto-van Silfhout, T. Nakagawa, N. Bahi-Buisson, S. A. Haas, H. Hu, M. Bienek, L. E. L.  
1219 M. Vissers, C. Gilissen, A. Tzschach, A. Busche, J. Müsebeck, P. Rump, I. B. Mathijssen, K. Avela,  
1220 M. Somer, F. Doagu, A. K. Philips, A. Rauch, A. Baumer, K. Voesenek, K. Poirier, J. Vigneron, D.  
1221 Amram, S. Odent, M. Nawara, E. Obersztyń, J. Lenart, A. Charzewska, N. Lebrun, U. Fischer, W.  
1222 M. Nillesen, H. G. Yntema, I. Järvelä, H.-H. Ropers, B. B. A. de Vries, H. G. Brunner, H. van  
1223 Bokhoven, F. L. Raymond, M. A. A. P. Willemsen, J. Chelly, Y. Xiong, A. J. Barkovich, V. M.  
1224 Kalscheuer, T. Kleefstra, A. P. M. de Brouwer, Variants in CUL4B are associated with cerebral  
1225 malformations. *Hum. Mutat.* **36**, 106–117 (2015).
- 1226 232. M. Co, A. G. Anderson, G. Konopka, FOXP transcription factors in vertebrate brain  
1227 development, function, and disorders. *Wiley Interdiscip. Rev. Dev. Biol.* **9**, e375 (2020).



- 1228 233. H. Ito, R. Morishita, M. Mizuno, N. Kawamura, H. Tabata, K.-I. Nagata, Biochemical and  
1229 Morphological Characterization of a Neurodevelopmental Disorder-Related Mono-ADP-  
1230 Ribosylhydrolase, MACRO Domain Containing 2. *Dev. Neurosci.* **40**, 278-287 (2018).
- 1231 234. H. Hor, L. Francescato, L. Bartesaghi, S. Ortega-Cubero, M. Kousi, O. Lorenzo-Betancor, F. J.  
1232 Jiménez-Jiménez, A. Gironell, J. Clarimón, O. Drechsel, J. A. G. Agúndez, D. Kenzelmann Broz, R.  
1233 Chiquet-Ehrismann, A. Lleó, F. Coria, E. García-Martin, H. Alonso-Navarro, M. J. Martí, J.  
1234 Kulisevsky, C. N. Hor, S. Ossowski, R. Chrast, N. Katsanis, P. Pastor, X. Estivill, Missense  
1235 mutations in TENM4, a regulator of axon guidance and central myelination, cause essential tremor.  
1236 *Hum. Mol. Genet.* **24**, 5677–5686 (2015).
- 1237 235. N. Hamajima, M. Kouwaki, P. Vreken, K. Matsuda, S. Sumi, M. Imaeda, S. Ohba, K. Kidouchi,  
1238 M. Nonaka, M. Sasaki, N. Tamaki, Y. Endo, R. De Abreu, J. Rotteveel, A. van Kuilenburg, A. van  
1239 Gennip, H. Togari, Y. Wada, Dihydropyrimidinase deficiency: structural organization, chromosomal  
1240 localization, and mutation analysis of the human dihydropyrimidinase gene. *Am. J. Hum. Genet.* **63**,  
1241 717–726 (1998).
- 1242 236. S. Mechaussier, B. Almoallem, C. Zeitz, K. Van Schil, L. Jeddawi, J. Van Dorpe, A. Dueñas Rey,  
1243 C. Condroyer, O. Pelle, M. Polak, N. Boddaert, N. Bahi-Buisson, M. Cavallin, J.-L. Bacquet, A.  
1244 Mouallem-Bézière, O. Zambrowski, J. A. Sahel, I. Audo, J. Kaplan, J.-M. Rozet, E. De Baere, I.  
1245 Perrault, Loss of Function of RIMS2 Causes a Syndromic Congenital Cone-Rod Synaptic Disease  
1246 with Neurodevelopmental and Pancreatic Involvement. *Am. J. Hum. Genet.* **106**, 859–871 (2020).
- 1247 237. P.-J. Mei, J. Bai, F.-A. Miao, C. Chen, Y.-S. Zhu, Z.-L. Li, J.-N. Zheng, Y.-C. Fan, CTHRC1  
1248 mediates multiple pathways regulating cell invasion, migration and adhesion in glioma. *Int. J. Clin.*  
1249 *Exp. Pathol.* **10**, 9318–9329 (2017).
- 1250 238. J. Nousbeck, R. Spiegel, A. Ishida-Yamamoto, M. Indelman, A. Shani-Adir, N. Adir, E. Lipkin,

- 1251 S. Bercovici, D. Geiger, M. A. van Steensel, P. M. Steijlen, R. Bergman, A. Bindereif, M. Choder, S.  
1252 Shalev, E. Sprecher, Alopecia, neurological defects, and endocrinopathy syndrome caused by  
1253 decreased expression of RBM28, a nucleolar protein associated with ribosome biogenesis. *Am. J.*  
1254 *Hum. Genet.* **82**, 1114–1121 (2008).
- 1255 239. S. R. F. Twigg, R. B. Hufnagel, K. A. Miller, Y. Zhou, S. J. McGowan, J. Taylor, J. Craft, J. C.  
1256 Taylor, S. L. Santoro, T. Huang, R. J. Hopkin, A. F. Brady, J. Clayton-Smith, C. L. Clericuzio, D. K.  
1257 Grange, L. Groesser, C. Hafner, D. Horn, I. K. Temple, W. B. Dobyns, C. J. Curry, M. C. Jones, A.  
1258 O. M. Wilkie, A Recurrent Mosaic Mutation in SMO, Encoding the Hedgehog Signal Transducer  
1259 Smoothened, Is the Major Cause of Curry-Jones Syndrome. *Am. J. Hum. Genet.* **98**, 1256–1265  
1260 (2016).
- 1261 240. T.-L. Le, Y. Sribudiani, X. Dong, C. Huber, C. Kois, G. Baujat, C. T. Gordon, V. Mayne, L.  
1262 Galmiche, V. Serre, N. Goudin, M. Zarhrate, C. Bole-Feysot, C. Masson, P. Nitschké, F. W.  
1263 Verheijen, L. Pais, A. Pelet, S. Sadedin, J. A. Pugh, N. Shur, S. M. White, S. El Chehadeh, J.  
1264 Christodoulou, V. Cormier-Daire, R. M. W. Hofstra, S. Lyonnet, T. Y. Tan, T. Attié-Bitach, W. S.  
1265 Kerstjens-Frederikse, J. Amiel, S. Thomas, Bi-allelic Variations of SMO in Humans Cause a Broad  
1266 Spectrum of Developmental Anomalies Due to Abnormal Hedgehog Signaling. *Am. J. Hum. Genet.*  
1267 **106**, 779–792 (2020).
- 1268 241. O. V. Chechneva, F. Mayrhofer, D. J. Daugherty, R. G. Krishnamurty, P. Bannerman, D. E.  
1269 Pleasure, W. Deng, A Smoothened receptor agonist is neuroprotective and promotes regeneration  
1270 after ischemic brain injury. *Cell Death Dis.* **5**, e1481 (2014).
- 1271 242. X. Fan, V. P. Masamsetti, J. Q. Sun, K. Engholm-Keller, P. Osteil, J. Studdert, M. E. Graham, N.  
1272 Fossat, P. P. Tam, TWIST1 and chromatin regulatory proteins interact to guide neural crest cell  
1273 differentiation. *Elife.* **10**, e62873 (2021).

- 1274 243. M. Yu, L. Ma, Y. Yuan, X. Ye, A. Montagne, J. He, T.-V. Ho, Y. Wu, Z. Zhao, N. Sta Maria, R.  
1275 Jacobs, M. Urata, H. Wang, B. V. Zlokovic, J.-F. Chen, Y. Chai, Cranial Suture Regeneration  
1276 Mitigates Skull and Neurocognitive Defects in Craniosynostosis. *Cell*. **184**, 243–256.e18 (2021).
- 1277 244. S. A. Mikheeva, A. M. Mikheev, A. Petit, R. Beyer, R. G. Oxford, L. Khorasani, J.-P. Maxwell,  
1278 C. A. Glackin, H. Wakimoto, I. González-Herrero, I. Sánchez-García, J. R. Silber, P. J. Horner, R. C.  
1279 Rostomily, TWIST1 promotes invasion through mesenchymal change in human glioblastoma. *Mol.*  
1280 *Cancer*. **9**, 194 (2010).
- 1281 245. S. Q. Shen, C. A. Myers, A. E. Hughes, L. C. Byrne, J. G. Flannery, J. C. Corbo, Massively  
1282 parallel cis-regulatory analysis in the mammalian central nervous system. *Genome Res*. **26**, 238–255  
1283 (2016).
- 1284 246. S. Zhang, J. Li, R. Lea, K. Vleminckx, E. Amaya, Fezf2 promotes neuronal differentiation  
1285 through localised activation of Wnt/ $\beta$ -catenin signalling during forebrain development.  
1286 *Development*. **141**, 4794–4805 (2014).
- 1287 247. O. Mercati, A. Danckaert, G. André-Leroux, M. Bellinzoni, L. Gouder, K. Watanabe, Y.  
1288 Shimoda, R. Grailhe, F. De Chaumont, T. Bourgeron, I. Cloëz-Tayarani, Contactin 4, -5 and -6  
1289 differentially regulate neuritogenesis while they display identical PTPRG binding sites. *Biol. Open*.  
1290 **2**, 324–334 (2013).
- 1291 248. C. Badouel, M. A. Zander, N. Liscio, M. Bagherie-Lachidan, R. Sopko, E. Coyaud, B. Raught, F.  
1292 D. Miller, H. McNeill, Fat1 interacts with Fat4 to regulate neural tube closure, neural progenitor  
1293 proliferation and apical constriction during mouse brain development. *Development*. **142**, 2781–2791  
1294 (2015).
- 1295 249. Q. Zhang, X. Gao, C. Li, C. Feliciano, D. Wang, D. Zhou, Y. Mei, P. Monteiro, M. Anand, S.  
1296 Itohara, X. Dong, Z. Fu, G. Feng, Impaired Dendritic Development and Memory in Sorbs2 Knock-

- 1297 Out Mice. *J. Neurosci.* **36**, 2247–2260 (2016).
- 1298 250. J. D. Ortigoza-Escobar, A. Oyarzabal, R. Montero, R. Artuch, C. Jou, C. Jiménez, L. Gort, P.  
1299 Briones, J. Muchart, E. López-Gallardo, S. Emperador, E. R. Pesini, J. Montoya, B. Pérez, P.  
1300 Rodríguez-Pombo, B. Pérez-Dueñas, Ndufs4 related Leigh syndrome: A case report and review of  
1301 the literature. *Mitochondrion.* **28**, 73–78 (2016).
- 1302 251. A. González-Quintana, M. J. Trujillo-Tiebas, A. L. Fernández-Perrone, A. Blázquez, A. Lucia,  
1303 M. Morán, C. Ugalde, J. Arenas, C. Ayuso, M. A. Martín, Uniparental isodisomy as a cause of  
1304 mitochondrial complex I respiratory chain disorder due to a novel splicing NDUFS4 mutation. *Mol.*  
1305 *Genet. Metab.* **131**, 341–348 (2020).
- 1306 252. E. Vantroys, A. Larson, M. Friederich, K. Knight, M. A. Swanson, C. A. Powell, J. Smet, S.  
1307 Vergult, B. De Paepe, S. Seneca, H. Roeyers, B. Menten, M. Minczuk, A. Vanlander, J. Van Hove,  
1308 R. Van Coster, New insights into the phenotype of FARS2 deficiency. *Mol. Genet. Metab.* **122**, 172–  
1309 181 (2017).
- 1310 253. R. Qin, S. Cao, T. Lyu, C. Qi, W. Zhang, Y. Wang, CDYL Deficiency Disrupts Neuronal  
1311 Migration and Increases Susceptibility to Epilepsy. *Cell Rep.* **18**, 380–390 (2017).
- 1312 254. A. Zito, D. Cartelli, G. Cappelletti, A. Cariboni, W. Andrews, J. Parnavelas, A. Poletti, M.  
1313 Galbiati, Neuritin 1 promotes neuronal migration. *Brain Struct. Funct.* **219**, 105–118 (2014).
- 1314 255. M. A. Kurian, E. Meyer, G. Vassallo, N. V. Morgan, N. Prakash, S. Pasha, N. A. Hai, S. Shuib, F.  
1315 Rahman, E. Wassmer, J. H. Cross, F. J. O’Callaghan, J. P. Osborne, I. E. Scheffer, P. Gissen, E. R.  
1316 Maher, Phospholipase C beta 1 deficiency is associated with early-onset epileptic encephalopathy.  
1317 *Brain.* **133**, 2964–2970 (2010).
- 1318 256. Y. R. Yang, D.-S. Kang, C. Lee, H. Seok, M. Y. Follo, L. Cocco, P.-G. Suh, Primary  
1319 phospholipase C and brain disorders. *Adv. Biol. Regul.* **61**, 80–85 (2016).

- 1320 257. R. Sueda, I. Imayoshi, Y. Harima, R. Kageyama, High Hes1 expression and resultant Ascl1  
1321 suppression regulate quiescent vs. active neural stem cells in the adult mouse brain. *Genes Dev.* **33**,  
1322 511–523 (2019).
- 1323 258. T. Matsuzaki, T. Yoshihara, T. Ohtsuka, R. Kageyama, Hes1 expression in mature neurons in the  
1324 adult mouse brain is required for normal behaviors. *Sci. Rep.* **9**, 8251 (2019).
- 1325 259. L. El-Bazzal, A. Atkinson, A.-C. Gillart, M. Obeid, V. Delague, A. Mégarbané, A novel EXT2  
1326 mutation in a consanguineous family with severe developmental delay, microcephaly, seizures,  
1327 feeding difficulties, and osteopenia extends the phenotypic spectrum of autosomal recessive EXT2-  
1328 related syndrome (AREXT2). *Eur. J. Med. Genet.* **62**, 259–264 (2019).
- 1329 260. A. Gupta, S. A. Ewing, D. L. Renaud, L. Hasadsri, K. M. Raymond, E. W. Klee, R. H. Gavriloza,  
1330 Developmental delay, coarse facial features, and epilepsy in a patient with EXT2 gene variants. *Clin*  
1331 *Case Rep.* **7**, 632–637 (2019).
- 1332 261. O. B. Shevelev, V. I. Rykova, L. A. Fedoseeva, E. Y. Leberfarb, G. M. Dymshits, N. G.  
1333 Kolosova, Expression of Ext1, Ext2, and heparanase genes in brain of senescence-accelerated OXYS  
1334 rats in early ontogenesis and during development of neurodegenerative changes. *Biochemistry* . **77**,  
1335 56–61 (2012).
- 1336 262. S. Hussain, Umm-E-Kalsoom, I. Ullah, K. Liaqat, S. Nawaz, W. Ahmad, A Novel Missense  
1337 Variant in the ALX4 Gene Underlies Mild to Severe Frontonasal Dysplasia in a Consanguineous  
1338 Family. *Genet. Test. Mol. Biomarkers.* **24**, 217–223 (2020).
- 1339 263. M. Kojic, T. Gawda, M. Gaik, A. Begg, A. Salerno-Kochan, N. D. Kurniawan, A. Jones, K.  
1340 Drożdżyk, A. Kościelniak, A. Chramiec-Głąbik, S. Hediye-Zadeh, M. Kasherman, W. J. Shim, E.  
1341 Sinniah, L. A. Genovesi, R. K. Abrahamsen, C. D. Fenger, C. G. Madsen, J. S. Cohen, A. Fatemi, Z.  
1342 Stark, S. Lunke, J. Lee, J. K. Hansen, M. F. Boxill, B. Keren, I. Marey, M. S. Saenz, K. Brown, S. A.

1343 Alexander, S. Mureev, A. Batzilla, M. J. Davis, M. Piper, M. Bodén, T. H. J. Burne, N. J. Palpant, R.  
1344 S. Møller, S. Glatt, B. J. Wainwright, Elp2 mutations perturb the epitranscriptome and lead to a  
1345 complex neurodevelopmental phenotype. *Nat. Commun.* **12**, 2678 (2021).

1346 264. Y.-T. Huang, J. O. Mason, D. J. Price, Lateral cortical *Cdca7* expression levels are regulated by  
1347 Pax6 and influence the production of intermediate progenitors. *BMC Neurosci.* **18**, 47 (2017).

1348 265. Y. Wei, X. Han, C. Zhao, PDK1 regulates the survival of the developing cortical interneurons.  
1349 *Mol. Brain.* **13**, 65 (2020).

1350

### 1351 Acknowledgements

1352 We would like to thank D. Genereux, D. Levesque, K. Lindblad-Toh, and the members of the Pfenning  
1353 Lab for useful discussions and suggestions. We would also like to thank P. Sullivan for curating the brain  
1354 size residual annotations, A. Hindle for providing us with annotations of which mammals spend time  
1355 underground, and M. Chikina, A. Kowalczyk, and E. Saputra for providing us early access to code for  
1356 phylogenetic permutations for phylogenetic generalized least squares. This work used the Extreme  
1357 Science and Engineering Discovery Environment (XSEDE), through the Pittsburgh Supercomputing  
1358 Center Bridges and Bridges-2 Compute Clusters, which was supported by National Science Foundation  
1359 grant number TG-BIO200055. Portions of this research were conducted on Lehigh University's Research  
1360 Computing infrastructure, which is partially supported by NSF Award 2019035.

1361 **Funding:** The Carnegie Mellon University Computational Biology Department Lane Fellowship  
1362 supported I.M.K. An NIH NIDA DP1DA046585 grant supported I.M.K., A.J.L., D.E.S., M.E.W., C.S.,  
1363 X.Z., B.N.P., A.R.B., and A.R.P. The Alfred P. Sloan Foundation Research Fellowship supported I.M.K.,  
1364 M.E.W., and A.R.P. An NSF Graduate Research Fellowship Program under grants DGE1252522 and  
1365 DGE1745016 supported A.J.L. A Carnegie Mellon University Summer Undergraduate Research  
1366 Fellowship supported D.E.S. An NIH NIDA Fellowship grant F30DA053020 supported B.N.P.

1367 **Author contributions:** I.M.K., A.J.L., and D.E.S. are listed as co-first authors in last name-alphabetical  
1368 order because they contributed equally to the manuscript. Conceptualization was done by I.M.K. and  
1369 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.  
1370 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  
1371 acquisition was done by A.R.P. with assistance from A.J.L. and B.N.P. Investigation was done by I.M.K.,  
1372 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  
1373 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  
1374 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.,  
1375 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.  
1376 and D.E.S. with assistance from C.S., A.J.L., and A.R.P. Manuscript preparation was done by I.M.K.,  
1377 D.E.S., and A.J.L. with assistance from A.R.P. and C.S. Manuscript review and editing was done by all  
1378 authors.

1379 **Competing interests:** No authors have any competing interests.

1380 **Data and materials availability:** Publicly available ATAC-seq data was obtained from Gene Expression  
1381 Omnibus accessions GSE161374, GSE146897, and GSE137311; China National GeneBank accession  
1382 CNP0000198; and ArrayExpress accession E-MTAB-2633. Unpublished ATAC-seq data generated by  
1383 the Pfenning Lab can be found at GSE159815 with token gzkbqwyobvmvfd and will be released prior to  
1384 publication. The tree used for the phenotype association pipeline can be obtained by contacting the  
1385 Zoonomia Consortium and will be released prior to publication. Publicly available genomes and  
1386 annotations were downloaded from NCBI Assembly and the UCSC Genome Browser. Publicly available  
1387 Hi-C data was downloaded from <http://hugin2.genetics.unc.edu/Project/hugin/>. Motif discovery results  
1388 and machine learning models can be found at  
1389 <http://daphne.compbio.cs.cmu.edu/files/ikaplow/TACITSupplement/>. Machine learning model predictions  
1390 can be obtained from authors upon request and will be released prior to publication. Code also used in our  
1391 previous work (25) can be found at <https://github.com/pfenninglab/OCROrthologPrediction>, new code for  
1392 this work can be found at <https://github.com/pfenninglab/TACIT>, and additional code can be obtained

1393 from authors upon request.

1394

1395 ZONOMIA CONSORITUM MEMBERS

1396 Gregory Andrews<sup>1</sup>, Joel C. Armstrong<sup>2</sup>, Matteo Bianchi<sup>3</sup>, Bruce W. Birren<sup>4</sup>, Kevin R Bredemeyer<sup>5</sup>, Ana  
1397 M Breit<sup>6</sup>, Matthew J Christmas<sup>3</sup>, Joana Damas<sup>7</sup>, Mark Diekhans<sup>2</sup>, Michael X. Dong<sup>3</sup>, Eduardo Eizirik<sup>8</sup>,  
1398 Kaili Fan<sup>1</sup>, Cornelia Fanter<sup>9</sup>, Nicole M. Foley<sup>5</sup>, Karin Forsberg-Nilsson<sup>10</sup>, Carlos J. Garcia<sup>11</sup>, John  
1399 Gatesy<sup>12</sup>, Steven Gazal<sup>13</sup>, Diane P. Genereux<sup>4</sup>, Daniel Goodman<sup>14</sup>, Linda Goodman<sup>15</sup>, Jenna Grimshaw<sup>11</sup>,  
1400 Michaela K. Halsey<sup>11</sup>, Andrew J Harris<sup>5</sup>, Glenn Hickey<sup>16</sup>, Michael Hiller<sup>17</sup>, Allyson Hindle<sup>9</sup>, Robert M.  
1401 Hubley<sup>18</sup>, Graham Hughes<sup>19</sup>, Jeremy Johnson<sup>4</sup>, David Juan<sup>20</sup>, Irene M. Kaplow<sup>21,22</sup>, Elinor K. Karlsson<sup>1,4</sup>,  
1402 Kathleen C. Keough<sup>23,24</sup>, Bogdan Kirilenko<sup>17</sup>, Jennifer M. Korstian<sup>11</sup>, Sergey V. Kozyrev<sup>3</sup>, Alyssa J.  
1403 Lawler<sup>25</sup>, Colleen Lawless<sup>19</sup>, Danielle L. Levesque<sup>6</sup>, Harris A. Lewin<sup>7,26,27</sup>, Xue Li<sup>1,4</sup>, Abigail Lind<sup>23,24</sup>,  
1404 Kerstin Lindblad-Toh<sup>3,4</sup>, Voichita D. Marinescu<sup>3</sup>, Tomas Marques-Bonet<sup>20,28,29,30</sup>, Victor C Mason<sup>31</sup>,  
1405 Jennifer R. S. Meadows<sup>3</sup>, Jill Moore<sup>1</sup>, Diana D. Moreno-Santillan<sup>11</sup>, Kathleen M. Morrill<sup>1,4</sup>, Gerard  
1406 Muntane<sup>20</sup>, William J Murphy<sup>5</sup>, Arcadi Navarro<sup>20,32,33,34</sup>, Martin Nweeia<sup>35,36,37,38</sup>, Austin Osmanski<sup>11</sup>,  
1407 Benedict Paten<sup>2</sup>, Nicole S. Paulat<sup>11</sup>, Eric Pederson<sup>3</sup>, Andreas R. Pfenning<sup>21,22</sup>, BaDoi N. Phan<sup>21</sup>, Katherine  
1408 S. Pollard<sup>23,24,39</sup>, Kavya Prasad<sup>21</sup>, Henry Pratt<sup>1</sup>, David A. Ray<sup>11</sup>, Jeb Rosen<sup>18</sup>, Irina Ruf<sup>40</sup>, Louise Ryan<sup>19</sup>,  
1409 Oliver A. Ryder<sup>41,42</sup>, Daniel Schäffer<sup>21</sup>, Aitor Serres<sup>20</sup>, Beth Shapiro<sup>43,44</sup>, Arian F. A. Smit<sup>18</sup>, Mark  
1410 Springer<sup>45</sup>, Chaitanya Srinivasan<sup>21</sup>, Cynthia Steiner<sup>46</sup>, Jessica M. Storer<sup>18</sup>, Patrick F. Sullivan<sup>47,48</sup>, Kevin  
1411 A. M. Sullivan<sup>10</sup>, Elisabeth Sundström<sup>3</sup>, Megan A Supple<sup>44</sup>, Ross Swofford<sup>4</sup>, Joy-El Talbot<sup>49</sup>, Emma  
1412 Teeling<sup>19</sup>, Jason Turner-Maier<sup>4</sup>, Alejandro Valenzuela<sup>20</sup>, Franziska Wagner<sup>40</sup>, Ola Wallerman<sup>3</sup>, Chao  
1413 Wang<sup>3</sup>, Juehan Wang<sup>13</sup>, Zhiping Weng<sup>1</sup>, Aryn P. Wilder<sup>41</sup>, Morgan E. Wirthlin<sup>21,22</sup>, Shuyang Yao<sup>48</sup>,  
1414 Xiaomeng Zhang<sup>21</sup>

1415

1416 1 Program in Bioinformatics and Integrative Biology, University of Massachusetts Medical School,  
1417 Worcester, MA 01605, USA

1418 2 Genomics Institute, UC Santa Cruz, 1156 High Street, Santa Cruz, CA 95064, USA

1419 3 Science for Life Laboratory, Department of Medical Biochemistry and Microbiology,  
1420 Uppsala University, Uppsala, 751 32, Sweden

1421 4 Broad Institute of MIT and Harvard, Cambridge MA 02139, USA

1422 5 Veterinary Integrative Biosciences, Texas A&M University, College Station, TX 77843, USA

1423 6 School of Biology and Ecology, University of Maine, Orono, Maine 04469, USA

1424 7 The Genome Center, University of California Davis, Davis, CA 95616, USA

1425 8 School of Health and Life Sciences, Pontifical Catholic University of Rio Grande do Sul, Porto Alegre,  
1426 90619-900, Brazil

1427 9 School of Life Sciences, University of Nevada Las Vegas, Las Vegas, NV 89154, USA

1428 10 Department of Immunology, Genetics and Pathology, Science for Life Laboratory, Uppsala  
1429 University, Uppsala, 751 85, Sweden

1430 11 Department of Biological Sciences, Texas Tech University, Lubbock, TX 79409, USA

1431 12 Division of Vertebrate Zoology, American Museum of Natural History, New York, NY 10024, USA

1432 13 Keck School of Medicine, University of Southern California, Los Angeles, CA 90033, USA

1433 14 University of California San Francisco, San Francisco, CA 94143 USA

1434 15 Fauna Bio Inc., Emeryville, CA 94608, USA

1435 16 Baskin School of Engineering, University of California Santa Cruz, Santa Cruz, CA 95064, USA

1436 17 Max Planck Institute of Molecular Cell Biology and Genetics, 01307, Dresden, Germany

1437 18 Institute for Systems Biology, Seattle, WA 98109, USA



- 1438 19 School of Biology and Environmental Science, University College Dublin, Belfield, Dublin 4, Ireland  
1439 20 Institute of Evolutionary Biology (UPF-CSIC), Department of Experimental and Health  
1440 Sciences, Universitat Pompeu Fabra, Barcelona, 08003, Spain  
1441 21 Department of Computational Biology, School of Computer Science, Carnegie Mellon University,  
1442 Pittsburgh, PA 15213, USA  
1443 22 Neuroscience Institute, Carnegie Mellon University, Pittsburgh, PA 15213, USA  
1444 23 Gladstone Institutes, San Francisco, CA 94158, USA  
1445 24 Department of Epidemiology & Biostatistics, University of California, San Francisco, CA 94158, USA  
1446 25 Department of Biology, Carnegie Mellon University, Pittsburgh, PA 15213, USA  
1447 26 Department of Evolution and Ecology, University of California, Davis, CA 95616, USA  
1448 27 John Muir Institute for the Environment, University of California, Davis, CA 95616, USA  
1449 28 Catalan Institution of Research and Advanced Studies (ICREA), 08010, Barcelona, Spain  
1450 29 CNAG-CRG, Centre for Genomic Regulation, Barcelona Institute of Science and Technology (BIST),  
1451 08036, Barcelona, Spain  
1452 30 Institut Català de Paleontologia Miquel Crusafont, Universitat Autònoma de Barcelona, 08193,  
1453 Cerdanyola del Vallès, Barcelona, Spain  
1454 31 Institute of Cell Biology, University of Bern, 3012 Bern, Switzerland  
1455 32 Catalan Institution of Research and Advanced Studies (ICREA), 08010, Barcelona, Spain  
1456 33 CRG, Centre for Genomic Regulation, Barcelona Institute of Science and Technology (BIST), 08036,  
1457 Barcelona, Spain  
1458 34 BarcelonaBeta Brain Research Center, Pasqual Maragall Foundation, Barcelona, 08005 Spain  
1459 35 Narwhal Genome Initiative, Department of Restorative Dentistry and Biomaterials Sciences, Harvard  
1460 School of Dental Medicine, Boston, MA 02115, USA  
1461 36 Department of Comprehensive Care, School of Dental Medicine, Case Western Reserve University,  
1462 Cleveland, OH 44106, USA  
1463 37 Department of Vertebrate Zoology, Smithsonian Institution, Washington, DC 20002, USA  
1464 38 Department of Vertebrate Zoology, Canadian Museum of Nature, Ottawa, Ontario K2P 2R1, Canada  
1465 39 Chan Zuckerberg Biohub, San Francisco, CA 94158, USA  
1466 40 Division of Messel Research and Mammalogy, Senckenberg Research Institute and Natural History  
1467 Museum Frankfurt, 60325 Frankfurt am Main, Germany  
1468 41 Conservation Genetics, San Diego Zoo Wildlife Alliance, Escondido, CA 92027, USA  
1469 42 Department of Evolution, Behavior and Ecology, Division of Biology, University of California, San  
1470 Diego, La Jolla, CA 92039 USA  
1471 43 Howard Hughes Medical Institute, University of California Santa Cruz, Santa Cruz, CA 95064, USA  
1472 44 Department of Ecology and Evolutionary Biology, University of California Santa Cruz, Santa Cruz,  
1473 CA 95064, USA  
1474 45 Department of Evolution, Ecology and Organismal Biology, University of California, Riverside,  
1475 CA 92521, USA  
1476 46 Conservation Science Wildlife Health, San Diego Zoo Wildlife Alliance, Escondido CA 92027, USA  
1477 47 Department of Genetics, University of North Carolina Medical School, Chapel Hill, NC 27599, USA  
1478 48 Department of Medical Epidemiology and Biostatistics, Karolinska Institute, Stockholm, Sweden  
1479 49 Iris Data Solutions, LLC, Orono, ME 04473, USA