RESEARCH ARTICLE

Tempo and mode of gene expression evolution in the brain across primates

Katherine Rickelton[1,2]*, Trisha M Zintel[1,2], Jason Pizzollo[1,2], Emily Miller[1], John J Ely[3,4], Mary Ann Raghanti[5], William D Hopkins[6], Patrick R Hof[7,8], Chet C Sherwood[3], Amy L Bauernfeind[9,10], Courtney C Babbitt[1]*

[1]Department of Biology, University of Massachusetts Amherst, Amherst, United States; [2]Molecular and Cellular Biology Graduate Program, University of Massachusetts Amherst, Amherst, United States; [3]Department of Anthropology and Center for the Advanced Study of Human Paleobiology, The George Washington University, Washington, United States; [4]MAEBIOS Epidemiology Unit, Alamogordo, United States; [5]Department of Anthropology, School of Biomedical Sciences, and Brain Health Research Institute, Kent State University, Kent, United States; [6]Department of Comparative Medicine, Michale E. Keeling Center for Comparative Medicine,The University of Texas M D Anderson Cancer Centre, Bastrop, United States; [7]New York Consortium in Evolutionary Primatology, New York, United States; [8]Nash Family Department of Neuroscience and Friedman Brain Institute, Icahn School of Medicine at Mount Sinai, New York, United States; [9]Department of Neuroscience, Washington University School of Medicine, St. Louis, United States; [10]Department of Anthropology, Washington University in St. Louis, St. Louis, United States

Abstract Primate evolution has led to a remarkable diversity of behavioral specializations and pronounced brain size variation among species (Barton, 2012; DeCasien and Higham, 2019; Powell et al., 2017). Gene expression provides a promising opportunity for studying the molecular basis of brain evolution, but it has been explored in very few primate species to date (e.g. Khaitovich et al., 2005; Khrameeva et al., 2020; Ma et al., 2022; Somel et al., 2009). To understand the landscape of gene expression evolution across the primate lineage, we generated and analyzed RNA-seq data from four brain regions in an unprecedented eighteen species. Here, we show a remarkable level of variation in gene expression among hominid species, including humans and chimpanzees, despite their relatively recent divergence time from other primates. We found that individual genes display a wide range of expression dynamics across evolutionary time reflective of the diverse selection pressures acting on genes within primate brain tissue. Using our samples that represent a 190-fold difference in primate brain size, we identified genes with variation in expression most correlated with brain size. Our study extensively broadens the phylogenetic context of what is known about the molecular evolution of the brain across primates and identifies novel candidate genes for the study of genetic regulation of brain evolution.

Editor's evaluation

This is an important study that represents a significant contribution to our understanding of how gene expression in the primate brain has evolved across the extant primate phylogeny. It provides solid evidence for potential links between gene expression variation and brain size, although these are somewhat limited by the focus only on adult brains, since many key changes likely occur during development. Nevertheless, both the taxonomically broad data set and the analysis are likely to

***For correspondence:** krickelton@umass.edu (KR); cbabbitt@bio.umass.edu (CCB)

Competing interest: The authors declare that no competing interests exist.

Funding: See page 15

Preprinted: 21 April 2021
Received: 12 May 2021
Accepted: 25 January 2024
Published: 26 January 2024

Reviewing Editor: Jenny Tung, Duke University, United States

be of broad interest to the evolutionary biology, anthropology, and comparative neuroscience communities.

Introduction

Primates are distinguished from other mammals by their large brains relative to body size (***Boddy et al., 2012***; ***Martin, 1981***; ***Smaers et al., 2021***). Among the diversity of primate species, there is remarkable variation in behavioral specializations, including differences in social structure, spatial, dietary and visual ecology, and locomotion (***Powell et al., 2017***; ***Barton, 2012***; ***DeCasien and Higham, 2019***). Despite the impressive array of cognitive attributes displayed by primates, molecular and cellular studies investigating various aspects of brain evolution tend to sample from a small number of species to address questions of how humans are unique. In large part, the emphasis on human brain evolution is warranted. Humans are unmatched in possessing exceptionally large brains and unparalleled cognitive abilities, such as language (***Konopka and Roberts, 2016***; ***Rilling, 2014***). While valuable, the limited number of species included in prior research lacks a comprehensive perspective of the phylogenetic context in which the human brain evolved within the diversity of primates.

Although researchers have used a variety of approaches to assess whether the human brain is unique (***Stout and Hecht, 2017***) and how it might have evolved (***Hrvoj-Mihic et al., 2013***; ***Sousa et al., 2017a***), the full potential for using gene expression to evaluate patterns of brain evolution in primates has not yet been met. Upon observing the remarkable similarity between human and chimpanzee protein sequences, ***King and Wilson, 1975*** proposed that the basis of the physical and behavioral phenotypic differences between these two species must be found in changes within gene regulatory regions that drive expression. Previous studies have explored how changes in regulatory regions can influence gene expression but have often sampled various organs from species across broad spans of evolutionary time, such as mammals or vertebrates (***Brawand et al., 2011***; ***Breschi et al., 2016***). In studies focusing on gene expression in primate brain tissues, research has mostly focused on the neocortex and cerebellum in human, chimpanzee, and rhesus macaque (***Babbitt et al., 2010***; ***Blekhman et al., 2010***; ***Khaitovich et al., 2005***; ***Khaitovich et al., 2004***; ***Khrameeva et al., 2020***; ***Konopka et al., 2012***; ***Ma et al., 2022***; ***Somel et al., 2009***; ***Sousa et al., 2017b***). However, new insights can be gained by sampling at greater neuroanatomical resolution from a broader array of primates. Examining gene expression of the brain from a more comprehensive landscape empowers novel inquiry in primate brain evolution, including questions pertaining to the sources of variation that drive expression differences, rates of expression change across the primate phylogenetic tree, and genes that correlate with brain size across primates.

In the current study, we sampled prefrontal cortex (PFC), primary visual cortex (V1), hippocampus (HIP), and lateral cerebellum (CBL) from 18 primate species, the broadest diversity of primates sampled in any study of gene expression in the brain to date, including species from several rarely-studied lineages. Our dataset represents 70–90 million years (***Perelman et al., 2011***) of primate evolution, providing a more thorough understanding of the evolution of gene expression across primates and allowing for an unprecedented view of how gene expression in the brain has changed over time across all major clades of primate phylogeny.

Results

Most variation can be explained at the species level, not by brain region

To understand how gene expression in the brain has evolved across the primate lineage, we generated and analyzed RNA-seq data from 18 primate species (including five hominoids, four cercopithecoids, four platyrrhines, and five strepsirrhines, with 1–3 biological replicates) across four brain regions, including PFC, V1, HIP, and CBL (***Figure 1***, ***Supplementary file 1***). The transcriptomes and gene models were assembled de novo (***Haas et al., 2013***) (see Materials and methods). We quantified the expression of 15,017 orthologs within hominoids, and 3432 on-to-one orthologs across all 18 species (***Supplementary files 2 and 3***). Variation in interspecific mammalian gene expression has been shown to be less pronounced than that observed across samples from different organs,

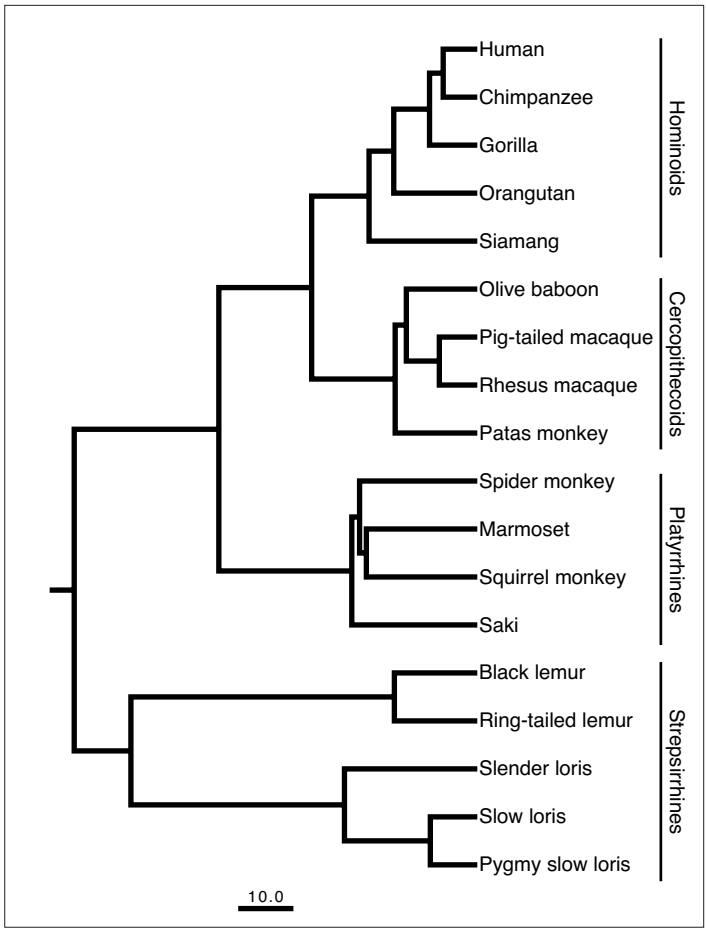

Figure 1. Primate phylogeny showing the eighteen species sampled in this study. The scale bar for the branch lengths represents 10 million years of evolution. The phylogenetic tree is a consensus tree of 1000 iterations produced from *10kTrees* v.3 (https://10ktrees.nunn-lab.org) based on data from GenBank. The insets demonstrate the approximate locations of the four brain regions sampled on a coronal section, midsagittal view, and lateral view (displayed left to right, respectively) of a schematized adult human brain.

reflecting the diversity of underlying organ physiology (***Brawand et al., 2011***; ***Sudmant et al., 2015***). Furthermore, it has been reported that the rate of gene expression divergence evolved more slowly within the cerebral cortex and cerebellum compared to other organ systems from developmentally distinct germ layers (***Brawand et al., 2011***; ***Khaitovich et al., 2005***). To explore the variability of gene expression from distinct regions of the brain across our broad sampling of primates, we constructed a pairwise distance matrix of the 500 most variable protein-coding genes based on the standard deviation of expression across samples (Methods). This subset of genes was enriched with glycoproteins, signal peptides, and plasma membrane proteins, with roles in immune function, molecular trafficking, and cell signaling. Using this distance matrix, we performed a principal coordinates analysis (PCoA) on data from all brain regions. Because our samples represent disparate regions of the same organ, we expected less variation to be attributed to brain regions than primate species or taxa, reflecting the similarity in physiology of brain tissues. Unsurprisingly, the variation from our complex gene expression dataset is represented across multiple axes of the PCoA (***Supplementary file 4***).

We plotted the first three axes and created polygons around data derived from samples sharing a common taxa (***Figure 2a–c***) or region (***Figure 2d–f***). As predicted, taxon assignment explains a large amount of variation to the dataset, with clear trends emerging independent of brain region. We find the greatest divergence in expression patterns among hominoid and strepsirrhine species, while there is more similarity observed among cercopithecoids and platyrrhines. The hominoids, displaying the greatest level of diversity of any primate phylogenetic group, demonstrate variation that is particularly apparent along Axis 1 and largely driven by human and chimpanzee expression patterns (***Figure 2A,***

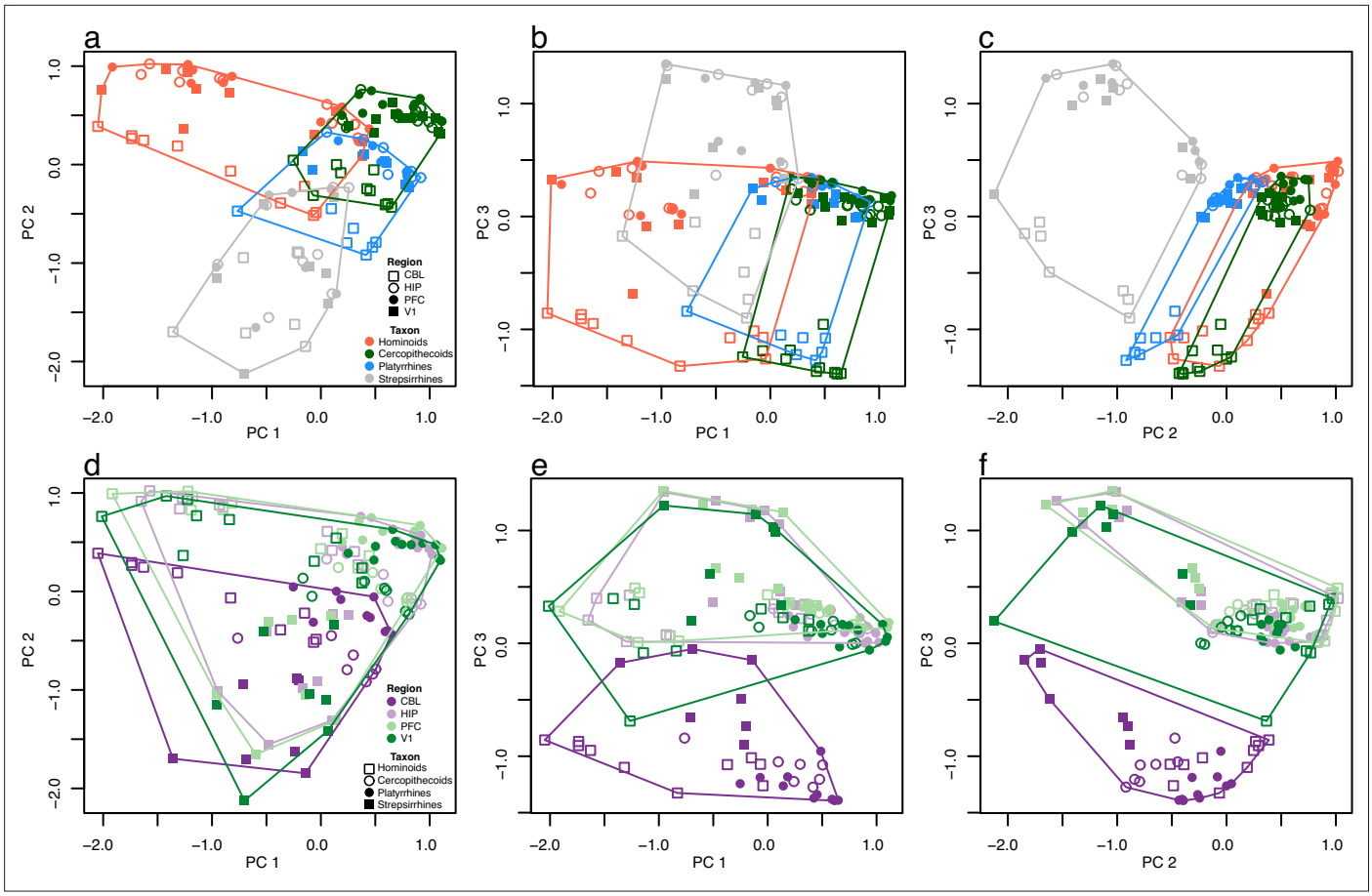

Figure 2. Patterns of brain gene expression across primates. The first three axes of a principal coordinates analysis (PCoA) are plotted in both rows but have different symbols and colors to emphasize expression patterns specific to taxa (upper row, **a–c**) and regions (lower row, **d-f**). Polygons in each plot surround the data points for taxa (upper row) and regions (lower row). Axes 1, 2, and 3 represent 12.8, 10.3, and 9.4% of variance, respectively.

The online version of this article includes the following figure supplement(s) for figure 2:

Figure supplement 1. The first three axes of the principal coordinates analysis (PCoA) are plotted in three bivariate plots.

Figure supplement 2. Gene expression phenogram of all sampled data.

Figure supplement 3. Gene expression phenograms by region.

B). Strepsirrhines also exhibit a large amount of variation, especially apparent along Axis 2, which can mostly be attributed to the three species of lorises. When Axis 1 and 2 of the PCoA are plotted on the same bivariate plot (***Figure 2a***), the hominoids display more variation than the strepsirrhines by about 24% (***Supplementary file 5***). However, a large portion of the variation in the strepsirrhines is attributed to evolutionary divergence over about 63 million years (since the last common ancestor of lemurs and lorises), whereas the variation within hominoids has largely accrued over only 9 million years (since humans and chimpanzees shared an ancestor with gorillas). The hominoid and strepsirrhine samples represent similar variation in terms of sex and life stage, suggesting that these factors do not account for the variability seen in these taxa. Therefore, a remarkable finding of this analysis is how much variation is represented by hominoids, despite the fact that this lineage represents a much shorter evolutionary divergence time.

The cerebellum differs most significantly from the other three sampled brain regions

We observed trends in gene expression by brain region that are predominantly seen along Axis 3 (***Figure 2e–f***). Here, we observe that the variation attributed to CBL is distinguished from that of PFC, V1, and HIP, which are very similar in their distributions. The fact that CBL differs in its pattern of gene

expression from other brain regions (***Hawrylycz et al., 2015***; ***Hawrylycz et al., 2012***; ***Itō, 2012***) is not surprising given that it is the only sampled region that develops from a different part of the embryonic neural tube (namely, the hindbrain vs. forebrain) and exhibits a neuronal packing density (predominantly glutamatergic granule cells) that far exceeds these other brain regions (***Azevedo et al., 2009***; ***Herculano-Houzel, 2011***; Itō, 2012). Enrichments for cerebellar gene expression reveal expression changes to categories such as 'Cell Surface Receptor Signaling Pathway,' 'Cell Projection Organization,' and 'Wnt Signaling Pathway.' This is true for humans in relation to chimpanzees as well as other species with deeper evolutionary relationships. (***Supplementary files 6 and 7***). Notably, genes involved in cell migration and cell surface receptor signaling ***Emera et al., 2016*** have been shown to mediate the cell-cell interactions necessary for axon guidance (***Koropouli and Kolodkin, 2014***). Although all the brain regions surveyed show some enrichment for categories related to cell signaling, the cerebellum shows a unique increase in signaling activity-related terms (both in the number of enrichment categories and the degree of change associated with these categories). This is indicative of region-specific differential expression.

Gene expression profiles accurately replicate phylogenetic relationships

To determine whether gene expression profiles can reconstruct known phylogenetic relationships among primates, we built expression phenograms (Methods) for all brain regions combined (***Figure 2—figure supplement 2***) and each region separately (***Figure 2—figure supplement 3***). Samples that were derived from individuals of the same species tended to be grouped together, regardless of brain region, revealing that inter-individual differences are minor compared to other sources of variation. Gene expression profiles also replicated the phylogenetic relationships of closely related species (e.g. humans and chimpanzees; pig-tailed and rhesus macaques) when all regions were considered, but these relationships became less phylogenetically structured in the phenograms constructed using expression data from individual brain regions. All neighbor-joining phenograms accurately represent cercopithecoids and strepsirrhines as monophyletic groups; however, expression data produces paraphyletic groups of hominoid and platyrrhine species. This result potentially reflects the fact that taxa with longer periods of independent evolution (i.e. strepsirrhines) are more likely to show divergent patterns of gene expression than more closely related groups. Meanwhile, a more dense sampling of cercopithecoids (three individuals per species), permits a fairly accurate reconstruction of this taxon.

Differential expression in the context of the Ornstein-Uhlenbeck model

Previous studies have used a variety of different approaches to model gene expression changes over time (***Brawand et al., 2011***; ***Perry et al., 2012***). Here, we used a recently described Ornstein–Uhlenbeck (OU) model to analyze neutral and conserved processes as determined by changing gene expression levels (***Chen et al., 2019***). OU processes have been proposed to model gene expression evolution as they model both drift and stabilizing selection (***Rohlfs et al., 2014***). Previous studies have shown how models that incorporate stabilizing selection are more accurately able to predict gene expression evolution in mammals than models that account only for neutral drift (***Chen et al., 2019***). Across all expressed one-to-one orthologs represented in the sampled primates, we found that ~15–20% of genes show differential expression across all species-to-species comparisons (q-value <0.05). As expected, the relative amount of differential expression increases over evolutionary time, both between species and clades (***Figure 3*** and ***Figure 3—figure supplement 1***). However, both the species and clade-wise comparisons show larger numbers of differentially expressed (DE) genes in the comparisons when strepsirrhines are included in the contrast, a result of the more than 60 million years of independent evolution of this taxon. In the prefrontal cortex, we see slightly smaller DE in humans as compared to siamang and baboon; however, in total number, this DE is not appreciably different (DE Human-Siamang n=251, DE Human-Baboon n=255) to other comparisons (DE Human-Rhesus Macaque n=315). Upset plots allow us to further analyze these unique patterns of differential expression (***Figure 3—figure supplement 2***). In the cerebellum, the lemur, baboon, and human samples are particularly unique and show relatively higher levels of DE compared to other species. However, in the hippocampus, visual cortex, and prefrontal cortex, the chimpanzees appear to show higher DE than the human samples. This suggests again that the cerebellum is a particularly unique brain structure

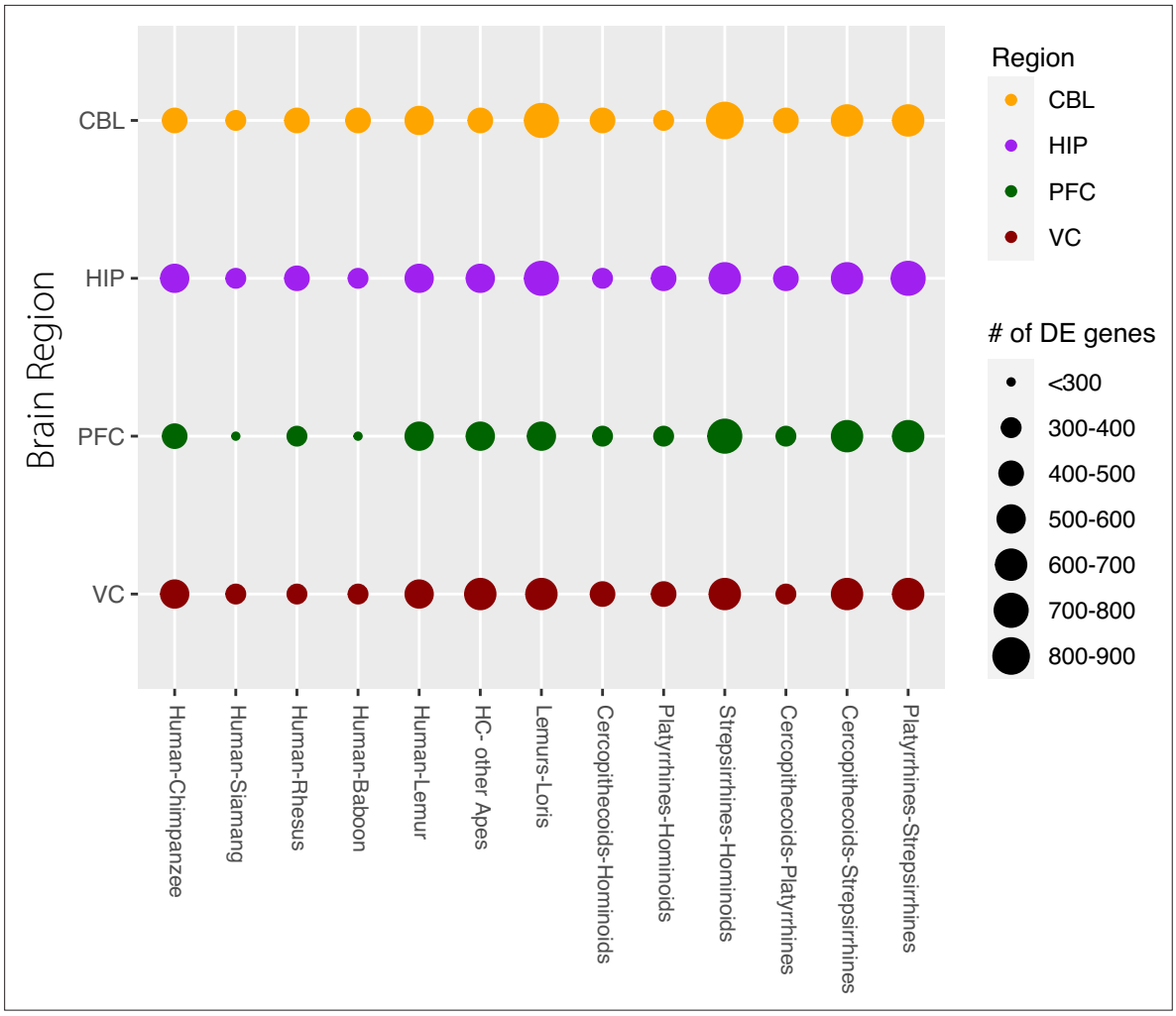

Figure 3. Gene counts of differentially expressed (DE) genes between species and clades. Each row represents one of the four brain regions examined. The size of the circle represents the number of DE genes seen at q<0.05 (5% FDR). The comparisons on the left are between exemplar species or sets of species, comparisons on the right are between clades of primates.

The online version of this article includes the following figure supplement(s) for figure 3:

Figure supplement 1. Box Plots of Pearson Rank-based Correlation Coefficients for multiple species and brain region comparisons.

Figure supplement 2. Upset Plot of each brain region (color) showing the shared DEGs for select species across the phylogeny.

and that chimpanzee gene expression is significantly different in the other three brain regions studied, warranting further analysis.

Human-specific enrichment for metabolic processes, neural development, and gene regulation

When comparing gene expression in the PFC of humans relative to other primates, human PFC shows an enrichment of metabolic processes, including 'regulation of cellular metabolic process' and 'regulation of macromolecule metabolic process' (***Supplementary files 6 and 7***). Comparing human and chimpanzee PFC reveals that categories that support neural growth and development (e.g. 'neuron projection morphogenesis,' 'cell morphogenesis involved in neuron differentiation'), gene regulation, and metabolic processes are enriched as differentially expressed in human PFC relative to chimpanzees.

In addition to examining human and chimpanzee data in isolation, we also analyzed human-specific changes in expression within the context of other outgroup species. Using the Ornstein-Uhlenbeck process as a model of continuous trait evolution across our 18-species primate phylogeny, we again

observe similar categories of enriched processes for the human prefrontal cortex in comparison to that of chimpanzees. These terms include 'Nervous System Development,' 'Neurogenesis,' 'Glial Cell Differentiation,' 'Neuron Projection Morphogenesis,' 'Regulation of Gene Expression,' and 'Metabolism.' To further validate our results, we also looked at the human PFC in comparison to other primate species. In analyzing differential expression between the human and siamang PFC, we note that similar trends for enrichment are also found, such as 'Neural Growth and Development' and 'Metabolic Processes,' and 'Gene Regulation.' Of interest, under the category of 'Positive Regulation of Transcription by RNA polymerase II' we find several genes that appear to be upregulated in humans compared to siamang: APP (amyloid precursor protein, related to plaque formation in Alzheimer's disease) as well as PRKN (found to be causal in Parkinson's disease) (***Funayama et al., 2023***). This supports the idea that these enrichments are human-specific, have relevance to important human neurodegenerative disease states, and are not a reflection of changes occurring within the chimpanzee lineage.

Broader species comparisons show similar trends across evolutionary time

Beyond human and chimpanzee comparisons, we also note many interesting broader temporal trends observed from the EVEE-based differential expression analysis. When examining Hominini (humans and chimpanzees) compared to other Great Apes, terms related to 'Regulation of Metabolic Process,' 'Nervous System Development,' and 'Biosynthetic processes' are all enriched within the hippocampus. Meanwhile, 'Negative Regulation of Synaptic Transmission' was enriched in the other ape species. In comparing the PFC of the Hominoid clade to that of the Strepsirrhine clade, we found that 'Neuron Development and Differentiation' was enriched. Overall, among various species and clade comparisons, there is a general trend of decreasing specificity in enrichment categories over increasing evolutionary time. Using PFC data, we found that the relationship between humans and chimpanzees, some of the closest relatives in our dataset, shows terms related to 'Synapse Assembly,' 'Regulation of Glial

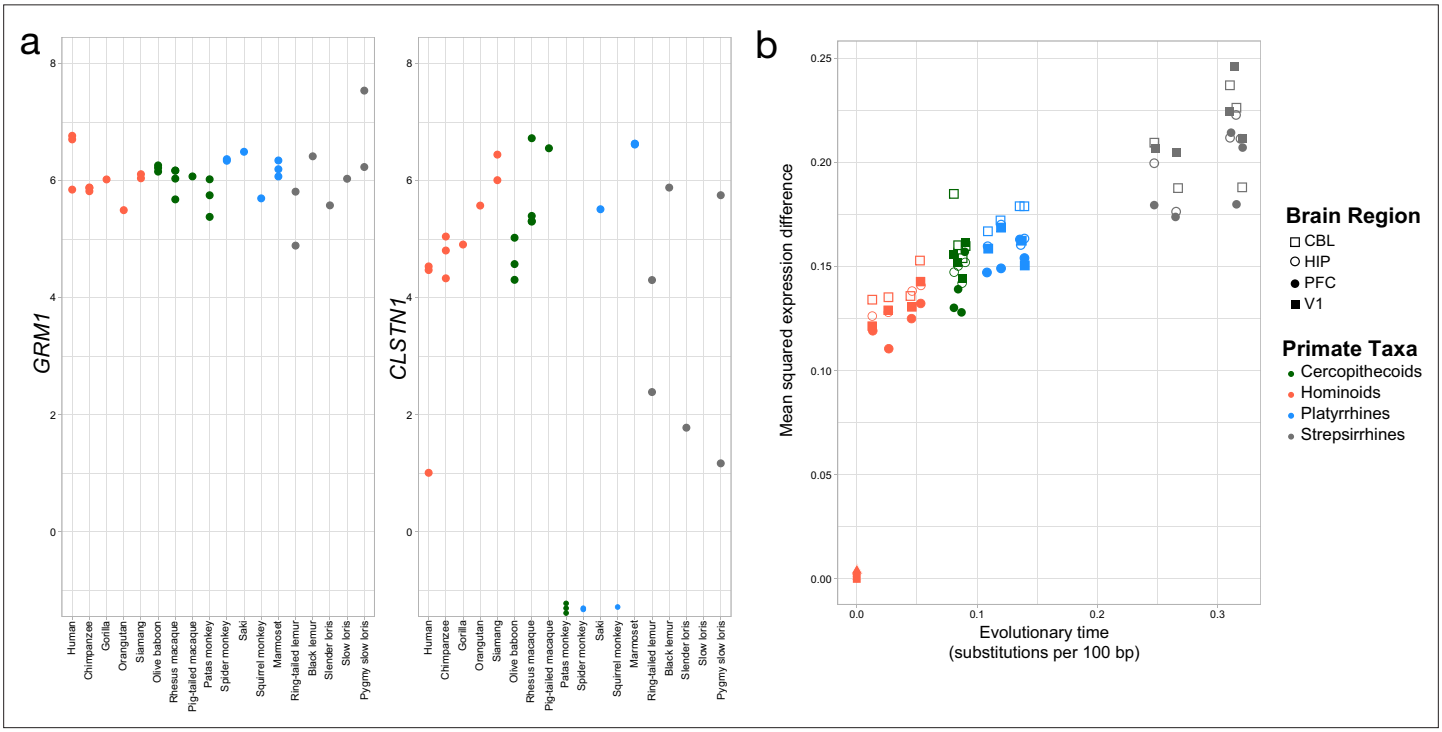

Figure 4. Rates of change over genes and evolutionary time. a. Exemplar genes that show constraint (left panel) and variation (right panel) across primates (colors as in ***Figure 2***). b. Mean squared expression difference plotted by evolutionary distance to humans across all orthologs that were expressed. Shapes denote the four brain regions, and the colors represent the four major primate clades represented in our samples.

The online version of this article includes the following figure supplement(s) for figure 4:

Figure supplement 1. Region-specific plots of mean square expression differences over evolutionary time for each of the four brain regions analyzed.

Cell Differentiation,' 'Regulation of Astrocyte Differentiation,' 'Axonogenesis,' and 'Neuron Projection Morphogenesis.' Looking at a more distantly related species pair, the human and rhesus macaque comparison shows enrichment for terms related to 'Cell Growth,' 'Cell Development,' 'Biological Regulation,' 'Neuron Projection Development,' 'Regulation of Neurogenesis,' and 'Positive Regulation of RNA Biosynthetic Processes.' The most distantly related species, humans and lemurs, have overall the largest number of differentially expressed genes, and with that, the broadest categories of enrichment. These include categories such as 'Regulation of Developmental Processes,' 'Regulation of Nervous System Development,' 'Cell Development,' and 'Multicellular Organism Development'.

Our PCoA analyses showed that gene expression in brain regions sampled from the lorises (i.e. the slender loris, slow loris, and pygmy slow loris) diverged from other strepsirrhines, and other primates more generally (***Figure 2—figure supplement 1***). When strepsirrhines are compared to other primates in differential expression analyses, transcription factors, and other genes involved in gene transcription and translation and multiple biosynthetic pathways involved in cellular metabolism are among the categories of DE genes (***Supplementary files 6 and 7***).

Evolutionary rates of expression change across clades and brain regions

Using the OU model, we found that individual genes exhibit wide variation in expression dynamics across the primate lineage (***Figure 4a***). Enrichments for genes showing low variation or stabilizing selection (q=0.05) reveal categories related to transport and cellular localization (GO Biological Processes, ***Supplementary files 6 and 7***). In contrast, genes that are less constrained or neutrally evolving (q>0.05) have a number of processes related to neuron morphogenesis, plasticity, and cell death. Yet, unlike sequence evolution, gene expression is not linear across evolutionary time but a saturation point in pairwise comparisons of gene expression is reached due to stabilizing selection pressures. Here, we find that pairwise expression differences between humans and the other species increasingly diverge with evolutionary distance in all brain regions sampled (***Figure 4b***); however, these pairwise comparisons do not seem to saturate with evolutionary time across the primate comparisons (***Chen et al., 2019***). We note that the saturation of pairwise expression differences from humans may be found at a phylogenetic node ancestral to primates (***Figure 3—figure supplement 1***).

Expression of a majority of genes evolves under stabilizing selection

We utilized EVEE-tools, developed in the context of the Ornstein-Uhlenbeck process of continuous trait evolution, to classify genes as primarily under the effects of stabilizing selection vs. a model of neutral drift (***Chen et al., 2019***). In this analysis, we found that across tissues, on average, 64% of genes fit better under stabilizing selection (64% CBL, 59% HIP, 72% PFC, and 60% V1; FDR-corrected q-value calculated via the BH procedure to correct for multiple hypothesis testing; FDR threshold of 5% to determine significance). In the context of the OU model of continuous trait evolution, we found that the overall phylogenetic signal in brain expression divergence was slightly smaller than observed with edgeR over the entire combined dataset (EVEE: 4–6.7% across all single species comparisons using a logFC of +/-2; edgeR: 15–20%). However, the relative amount of differential expression does increase gradually with evolutionary distance (as expected based on results in edgeR). The only exception to this is the human and chimpanzee comparison, which shows considerable variation in differential expression across the four brain regions. Representing such a short period of evolutionary divergence, this increase in DE suggests direction selection within that lineage.

Evidence of potential directional selection in human and chimpanzee data

To better understand the outlier gene expression in this dataset, and overall to gain insight into genes that may be subject to directional selection pressures, we again used the EVEE-tools OU-based model to score our dataset for outlier expression. This requires the determination of the evolutionary mean and variance for each gene across our entire expression dataset, from which we then can compare to individual species expression data. Using this method, we found a small subset of genes to be defined as having an expression that deviates from the optimal OU-distribution (Z-score >2 or < –2, p-value <0.05) (***Figure 4Figure 4—figure supplement 1***). It is important to note that a significant FDR (<5%) was not reached in this dataset, among all species comparisons. This is expected based on previous

applications of this package, in which a mammalian dataset was also unable to reach an FDR below 18% (***Chen et al., 2019***). This is likely a reflection of the limitations of our phylogeny, and suggests that future projects should aim to sample even more broadly.

In our outlier expression analysis, we found that patterns of outlier gene expression occur in a species and tissue-specific manner (***Supplementary file 10***). For example, in the PFC, the chimpanzee and marmoset samples appear to have the highest number of outlier genes deviating from average expression patterns (compared to human, siamang, baboon, rhesus macaque, and marmoset samples). We found this to be true regardless of how the dataset was normalized in order to determine the average expression for each gene (via defining a reference species). This is particularly interesting and, in combination with differential expression analysis, highlights the chimpanzee PFC as a particularly divergent structure. In contrast to the PFC, outlier analysis in the CBL reveals that the human and lemur samples have the highest number of outlier genes.

Upon gene set category enrichment analysis, we find that many of these genes that are deemed 'outliers' are related to functions in development, transcription, nervous system development, neurogenesis, and metabolism. For example, the chimpanzee PFC shows a significant upregulation in genes involved in energy storage and transfer, such as ETFA and NDUFS4, which are both involved in electron transport for ATP generation, as well as ANKH, implicated in phosphate transport (***Szeri et al., 2020***; ***Henriques et al., 2021***; ***Shil et al., 2021***) We also see that the chimpanzee PFC shows unique expression patterns in genes related to synaptic activity and neurotransmitter release, including significant downregulation of GABRA4 and SYN1 (***Fassio et al., 2011***; ***Fan et al., 2020 Fan et al., 2020***; ***Fassio et al., 2011***). In contrast, the human CBL and PFC both display a significant upregulation of genes related to Amyloid protein production (APP), a major component of many neurodegenerative diseases with functions in synaptic signaling (***O'Brien and Wong, 2011***). Unique to the human CBL we also see enrichments in genes related to neurogenesis and synaptic activity, including SDK1, FZD5, and CDH10 (***Redies et al., 2012***; ***Slater et al., 2013***; ***Bagot et al., 2016***) This is suggestive of directional selection pressures occurring in a tissue-specific manner and encourages future investigation of these outlier genes.

Implications of using humans as a reference species

We continue to use humans as a reference species in these analyses as, compared with other primates, humans have exceptionally large brain sizes and unique cognitive abilities. However, we do recognize that there are some implications for having humans as a reference, especially given our data that would suggest human gene expression as being largely different from the rest of our primate dataset. To address this, we repeated analyses using EVEE-tools by including two additional reference species: siamang and rhesus macaque. The percentage of genes that fit better under the model of stabilizing selection (in comparison to neutral drift) is not statistically different from those observed when using humans as a reference (on average, 58–64% across all four major brain regions). We additionally looked at pairwise species comparisons to determine if the general trends of directional selection and differential gene expression were comparable to the human-reference data, and again confirmed that the effects of the reference species used for normalization here are negligible in terms of analyzing differential expression across our primate tree.

Correlation of gene expression to brain size and their change over evolutionary time

Because an increase in absolute brain size is one of the most striking characteristics of humans, we asked what subset of genes is correlated with this important biological trait across the primate tree. Our dataset provides a unique opportunity to evaluate this question, since the average brain size across our study species varies by ~190 fold. Using multivariate analysis (Methods), we defined gene lists most strongly correlated with brain size in each brain region (***Supplementary file 8***). Results indicate that the same genes rank among those with the strongest positive correlation in PFC, V1, and HIP. CBL also shares some of these same genes but includes more variation among the genes most strongly correlated to overall brain size than the other three brain regions (***Figure 5***, ***Supplementary file 9***), potentially reflecting the more recent expansion of the CBL relative to the rest of the neocortex (***Miller et al., 2019***; ***Smaers et al., 2018***).

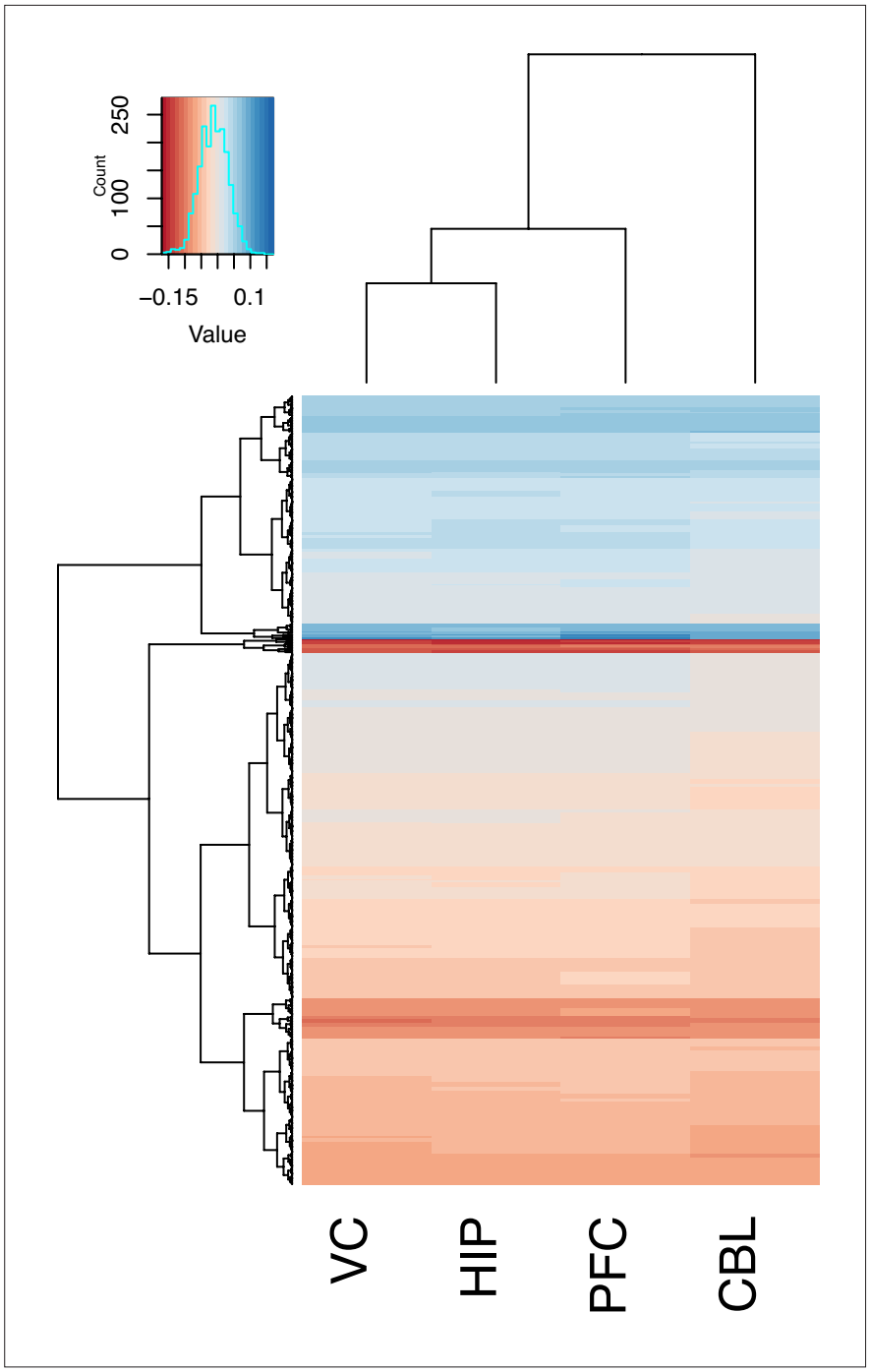

Figure 5. Gene correlated with brain size by region. A clustering and heatmap of the loadings from PC2 of genes for the four regions examined (V1, HIP, PFC, and CBL).

The online version of this article includes the following figure supplement(s) for figure 5:

Figure supplement 1. principal coordinates analysis (PCoA) of expression data from human samples of all four brain regions and primary neurons and astrocytes.

Discussion

We performed RNA-seq on four brain regions from 18 primate species, representing the broadest sampling for any gene expression study in primate brain tissue to date. Through more representative sampling of primate species, we found substantial variation in gene expression levels within the

hominoid and strepsirrhine lineages, with the diversity among hominoids particularly impressive due to the recent divergence of this taxon. Using the OU model, we found that a substantial proportion of genes showed differential expression across species. The relative amount of differential expression increased over evolutionary time, both between species and clades. Additionally, when comparing gene expression across broader species and clade comparisons, we observed trends related to brain development, nervous system regulation, and cellular metabolism.

Our findings point to human-specific enrichment for metabolic processes, neural development, and gene regulation. The considerable diversity of gene expression in human and chimpanzee brain tissue has profound implications for understanding the distinct evolutionary processes that have acted upon the brains of the ancestral species of these two lineages. We observed a wide variety of expression dynamics of individual genes in the pairwise comparisons of humans to other primate species. Enrichments for genes under stabilizing selection indicated processes related to transport and cellular localization, while less constrained genes were associated with neuron morphogenesis, plasticity, and cell death. Importantly, gene expression evolution did follow a linear pattern, but did not reach a saturation point due to stabilizing selection pressures as seen in other studies (***Chen et al., 2019***). Less constrained, neutrally evolving patterns appeared to be the most prevalent pattern in each brain region studied, preventing a saturation point of stabilizing pressures to be reached within Primates with increased phylogenetic distance from humans. Lastly, we identified genes that are correlated with brain size across all major primate taxa, providing candidates for further inquiry.

Our deeper analysis of gene expression has revealed evolutionary patterns that were inaccessible with a more limited sampling of primate brain tissue. We anticipate that the candidate genes and data provided by this study will serve as a resource for many other lines of inquiry into human and non-human primate brain evolution.

Materials and methods

Biological sample collection and RNA extraction

The sample includes brain tissue from human and nonhuman primates. All samples were obtained from adult individuals free from known neurological disease. If available, the right hemisphere was preferentially sampled. Human brain samples were obtained from the National Institute for Child Health and Human Development Brain and Tissue Bank for Developmental Disorders at the University of Maryland (Baltimore, MD). Chimpanzee brain tissue was obtained from the National Chimpanzee Brain Resource (https://www.chimpanzeebrain.org, supported by NIH grant NS092988). All other sources of brain tissue are listed in ***Supplementary file 1***.

From each individual, we sampled four regions of the brain, including PFC, V1, HIP, and CBL. PFC was sampled from the frontal pole, corresponding to Brodmann's area 10 in humans. In other primates, the PFC region sampled more broadly encompassed prefrontal cortical areas but was limited to the most anteriorly projecting part of the frontal pole. All V1 samples were dissected around the calcarine sulcus to include primary visual cortex (Brodmann's area 17). Samples from both PFC and V1 contained all cortical layers and a small amount of underlying white matter (<10%). The HIP was sampled from the medial aspect of the temporal lobe and included all hippocampal subfields. The CBL was sampled from the most laterally projecting region of lateral hemisphere in all primates. In humans, the CBL region corresponded to Crus I or Crus II. CBL samples contained all layers of cerebellar cortex and a small amount of underlying white matter (<10%). Each sample was briefly homogenized using a Tissuelyzer (Qiagen), and the total RNA was isolated using an RNAeasy kit (Qiagen) with a DnaseI treatment.

Library preparation and sequencing

Single-end RNA-seq libraries were made using the NEBNext mRNA Library Prep Reagent Set for Illumina. Libraries were prepared in batches of 4–8 samples of randomly sampled species and brain regions. Library sizes were checked on the Bioanalyzer (Agilent). RNA-seq libraries were multiplexed on the NextSeq500 (Illumina) in the Genomics Resource Laboratory at the University of Massachusetts Amherst, also randomly distributed across NextSeq500 runs. All fastq files have been submitted to the SRA: https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA639850.

Mapping and transcriptome analysis

Sequencing reads were assembled into species-specific transcriptomes (containing the reads from all four brain regions) using Trinity (***Grabherr et al., 2011***). With Trinity assembly, fragments of the original RNA reads for each species are compiled and clustered into groups based on sequence similarity, which eventually are extended to reconstruct full-length transcripts (***Grabherr et al., 2011***). These transcripts were then blasted against the human blast nt database, with the alignment thresholds for the top hits from different clades listed in ***Supplementary file 2***, similar to the approach in ***Perry et al., 2012***. Individual libraries were then mapped to the species-specific transcriptome using bowtie (***Langmead et al., 2009***) in RSEM, and count tables were generated using RSEM (***Li and Dewey, 2011***). Orthology assignments were additionally checked using the Ensemble one-to-one orthology alignments as a guide for the subset of species with a publicly available genome. Transcriptome quality was assessed using BUSCO to determine assembly and annotation completeness (***Seppey et al., 2019***), ***Supplementary file 3***. We recognize that some of the species transcriptomes show relatively lower BUSCO completeness scores (namely the Slender Loris at 35.8% complete). We hypothesize that this is likely a reflection of the limited tissue sampling in this dataset of only brain tissue. Previous studies have shown that transcriptome assemblies from single tissue regions on average have lower completeness scores than assemblies composed of reads from a variety of tissue types (***Simão et al., 2015***). This is likely a reflection of tissue-specific gene expression. As evidence of this, we further analyzed reads deemed as 'missing' by BUSCO and found that many of these showed little to no expression across the human brain (***Uhlén et al., 2015***) and Human Protein Atlas proteinatlas.org. We do not make any major conclusions about the Loris species in this manuscript and thus do not believe these BUSCO scores significantly affect the conclusions made in this manuscript. We only consider Loris data in concert with Lemurs, which by comparison have much more complete transcriptomes.

Distance-based data analyses (PCoA and phenograms)

We performed principal coordinates analyses (PCoA) based on a pairwise distance matrix of all 137 samples. The distance matrix was comprised of the top 500 most variably expressed protein-coding genes. Pairwise distances were calculated by leading log2 fold change, providing a symmetrical representation of the expression ratio centered around 0 (i.e. log2(2)=1 while log2(0.5) = –1) (***Robinson et al., 2010***). Creating the distance matrix and plotting the PCoA were performed using the plotMDS.DGElist function (based in limma) in the edgeR package in R. Although variation is represented across more than 20 axes (***Supplementary file 4***), the first three axes were plotted to compare patterns across primate taxa and brain region sampled (***Figure 2***, ***Figure 2—figure supplements 1–3***). Polygons overlap the data points representing taxa or brain region. The area of each polygon was computed using functions in the sp package of R. The chull() function was used to define the points around the perimeter of each polygon, and the Polygon() function calculated the area of each. Relative areas of each polygon are listed in ***Supplementary file 5***. ***Figure 3*** displays the same data as the PCoA but uses an array of colors allowing the data from each individual species to be visualized.

The same log2 fold change distance matrix was then used to create phenograms representing the similarity of gene expression profiles among samples. The minimum distance neighbor-joining function in the 'ape' package of R created a tree based on the method proposed by ***Saitou and Nei, 1987***. The boot.phylo function estimated the reliability of given nodes of the tree by resampling over 1000 iterations. Although our objective in this analysis was to investigate patterns of evolution across the primate order and the brain regions, our sample included multiple individuals from the same species. By treating these samples separately, our analyses represent both within- and between-species variability in gene expression over time.

Analyzing differential expression

Counts were filtered and normalized using edgeR (***Robinson et al., 2010***), with any multispecies comparisons using the GLM functionality (***McCarthy et al., 2012***). Gene Ontology enrichments were performed using the DAVID gene ontology tool 6.8 (***Huang et al., 2009a***; ***Huang et al., 2009b***) and g:Profiler (***Reimand et al., 2016***). ***Supplementary files 6 and 7*** shows results from ordered g:profiler enrichments (g:GOSt) performed on DE genes where q<0.05 (note, this is not ranked on polarity of expression, just absolute change) with all genes expressed in this study used as background.

Differential Expression was also analyzed using the package EVEE-tools to incorporate the Ornstein-Uhlenbeck model (***Chen et al., 2019***). Evolutionary means and variance values were calculated for each gene across the entire phylogeny for individual brain-region datasets as well as the entire RNA-seq bulk dataset in the context of the Ornstein-Uhlenbeck model. Differential expression was determined using a multivariate OU model. Regimes were defined by species/clades of interest (i.e. in human vs. chimpanzee comparisons, the human samples represent one regime in one model to compare to all other samples, represented by a second regime. In comparison, the chimpanzee samples would be represented as an additional model and a separate regime.) p-values were calculated to represent the fit of each OU model (i.e. Human or Chimp-specific expression patterns) against a Brownian motion model. These were then corrected for multiple hypothesis testing using the Benjamini-Hochberg FDR procedure. We used an FDR threshold of 0.05 to define significance. Additionally, Akaike and Bayesian Information Criterion (AIC and BIC) scores were calculated for each gene for each model, and only genes with AIC and BIC scores significant against the null were considered in further analysis. Directionality of differential expression was further determined by comparing estimated mean expression levels for each regime in each model.

We also looked at differential expression beyond pairwise comparisons, again using EVEE-tools, to see how each individual species differed across the entire dataset. With this, one species of interest was treated as a single regime while all 17 other species were grouped as a second regime. The same criteria as above were used to determine significant differential expression, and this data was utilized to construct UpSet plots using the UpsetR package (***Conway et al., 2017***). We included only relevant species in these UpSet plots (representing each major primate clade) to simplify the graphs. Each graph is also separated by brain region, similar to previous analyses (***Figure 3—figure supplement 2***).

We also analyzed sources of variation in our data to determine how significant of an effect species differences have compared to other factors, including primate families and individual variation. For this, we conducted a correlation analysis utilizing the non-parametric Mann-Whitney U (MWU) test on the differentially expressed genes between humans and chimpanzees to other primate species. We focused this analysis on only those species with three individuals per brain region (olive baboon, rhesus macaque, and lemurs). Using the set of human-chimpanzee DEGs, we calculated the Spearman rank-based correlation coefficient between each species to either human or chimpanzee expression. We determined whether or not these correlation coefficients were significantly different across species and across brain regions using the Mann-Whitney U test. We determined that there was no significant correlation between any of the three species to human or chimpanzee expression in any of the four brain regions (***Figure 3—figure supplement 1***; ***Yapar et al., 2021***). This suggests that the expression profiles of these more distantly related primates are equally similar to human and chimpanzee expression patterns. We also conducted additional Analysis of Covariates (ANCOVA) to confirm that other factors, namely age and sex, were not significant sources of variation in our gene expression analyses. These analyses, along with our PCoA plots, show that taxon identity and brain region are the two most significant determinants for DE. Additionally, differences in samples from the same individual are defined by differences at the level of brain region. ANCOVA analyses showed minor residual effects that we deemed as random and were not further analyzed for the purposes of this manuscript.

Outlier expression was determined using the scoreGenes.R script from the EVEE-Tools script suite (***Chen et al., 2019***). We compared the total dataset-normalized mean expression of a single gene to that of a single select species, in the context of the overall mean expression and variance across the dataset. We specifically analyzed outlier expression in a brain region-specific manner, looking at the datasets subset by individual brain region. To normalize the expression of the entire dataset, a single species was selected to be used as a reference in TMM normalization. For all outlier analyses except for humans, the human samples were used as a reference. For the human outlier analysis, the rhesus samples were used as a reference for normalization. Importantly, we tested the use of different reference species for dataset normalization and did not find a significant difference in the number of outlier genes and the enrichment categories associated. Z-scores were calculated for genes whose expression patterns fit an OU model (in comparison to the null model of Brownian Motion) and whose evolutionary mean is above 5 CPM for the entire dataset (***Chen et al., 2019***). In order to determine significant outliers, an FDR threshold of 0.05 was again employed, however at this level limited significance was found. ***Supplementary file 10*** shows the results of this analysis for the CBL and PFC brain regions.

Phylogenetic and evolutionary distance analysis

Categorical enrichments for the contrasts between species and clades are in ***Supplementary file 6***. The phylogenetic tree for primates was downloaded from the UCSC Genome Browser (30 primate species) (***Kent et al., 2002***). Distances between species were extracted using the Environment for Tree Exploration Toolkit (***Huerta-Cepas et al., 2010***). The residuals and mean squared expression differences of all orthologs across 18 species were found using the package EVEE (***Chen et al., 2019***), and in all contrasts, humans were used as the reference species. We then analyzed the subset of genes showing either broadly defined conserved (low) or neutral (higher) variation across species (low $<q = 0.05$, high $q>0.05$), with categorical enrichments for these two groups in ***Supplementary file 7***.

Comparisons of the heterogenous tissues used and single-cell gene expression data

In any study that derives results from homogenized tissue samples, the composition heterogeneity of the samples may drive differences in gene expression (***Montgomery and Mank, 2016***). To address this issue, we compared the expression of our tissue samples to recent studies that have performed single-cell RNA-seq on neurons and astrocytes. RNA-seq data from primary neurons and astrocytes were obtained from NCBI's Gene Expression Omnibus (GEO) and processed in the same manner as the tissue samples for all human samples. These included four hippocampal astrocytes, four cortical astrocytes, and one cortical neuron from ***Zhang et al., 2016*** (GEO accession number GSE73721) and three pyramidal neuron samples isolated from an unspecified brain region by the ENCODE project (***ENCODE Project Consortium, 2012***; ***Davis et al., 2018***) (accession numbers GSM2071331, GSM2071332, and GSM2071418). Only genes with counts greater than zero in all samples and (CPM)>1 in all 23 samples were included in this analysis (n=7111). A PCoA was made from a distance matrix of the top 500 most variably expressed genes by the pairwise biological coefficient of variation (method = "bcv") across samples (***Robinson et al., 2010***). Creating the distance matrix and plotting the PCoA were performed using the plotMDS.DGElistfunction in the edgeR package in R. The PCoA of our human samples in comparison to primary neurons and astrocytes suggests that our heterogeneous tissue samples are not biased to contain more neurons or astrocytes as compared to each other (i.e. one tissue is not biased within this small sample set) (***Figure 5—figure supplement 1***), and is consistent with other neural cell and brain tissue comparisons (***Khrameeva et al., 2020***).

Additionally, prior knowledge of variation across primates in cell type composition of the brain is informative in interpreting bulk RNA-seq data. It is well known that neuron densities tend to decrease as brain size increases (***Sherwood et al., 2020***). This suggests that larger brains accommodate a smaller number of neurons per unit volume compared to smaller brains. However, it is interesting to note that other structural elements, such as astrocytes (***Munger et al., 2022***), microglia (***Dos Santos et al., 2020***), and synapses (***Sherwood et al., 2020***), exhibit a relatively invariant density per unit volume across species. Despite changes in brain size, these essential components involved in neural communication and support maintain a consistent presence, emphasizing their crucial role in brain function regardless of the species' brain size. Thus, based on this previous research, in cross-species comparisons of bulk tissue, we can tentatively interpret differences in gene expression to reflect generally similar proportions of major cell types, except for neurons, which are expected to decrease in proportion with larger brain sizes.

Brain size analyses

Average species endocranial volumes (ECVs) were obtained primarily from ***Kamilar and Cooper, 2013*** and ***Isler et al., 2008*** from the mean of male and female volumes. ECVs were used since reliable brain size data does not exist for all species samples. The data for human ECV (also averaged from male and female data points) was previously published by ***Coqueugniot and Hublin, 2012***. ***Isler et al., 2008*** reported a minimal error when ECV was transformed to brain mass using the correction factor of 1.036 g/ml (***Stephan, 1960***), and we used this conversion to obtain brain mass estimates from ECVs for all species (***Supplementary file 8***).

Within each brain region, we performed a PCA on the species average gene expression of the 500 most variable genes by standard deviation, using the prcomp function in R. For each regional PCA, 14 PCs were required to account for about 90% of the variance in gene expression. We performed multiple regression analyses to determine which of the PCs could predict brain size. Using all 14 PCs

accounted for at least 95% of the variation in each brain region. Akaike information criterion was applied. However, it was noted that for each brain region, PC2 was the most predictive of brain size by low (regional adjusted R^2 values for PC2 against brain size were: PFC, 0.42; V1, 0.56; HIP, 0.50; CBL, 0.36). The 500 genes and their loadings on PC2 are listed in ***Supplementary file 9***. Across the four sampled regions, we find general uniformity in the extent to which individual genes affect brain size, but the cerebellum displays the most unique signature of the regions sampled (***Figure 5***).

Acknowledgements

We would also like to acknowledge our funding from NSF BCS-1750377, the Wenner Gren Foundation, James S McDonnell Foundation (220020293), NSF INSPIRE (SMA-1542848), NIH (NS-092988), and a fellowship to KR from NIH T32 GM135096. https://doi.org/10.37717/220020293.

Additional information

Funding

Funder	Grant reference number	Author
National Science Foundation	BCS-1750377	Courtney C Babbitt
National Institutes of Health	T32 GM135096	Katherine Rickelton
James S. McDonnell Foundation	https://doi.org/10.37717/220020293	Chet C Sherwood
National Institutes of Health	NS-092988	Chet C Sherwood
National Science Foundation	SMA-1542848	Chet C Sherwood

The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

Author contributions

Katherine Rickelton, Formal analysis, Visualization, Methodology, Writing – review and editing; Trisha M Zintel, Data curation, Formal analysis, Writing – review and editing; Jason Pizzollo, Formal analysis, Investigation, Methodology; Emily Miller, Formal analysis, Visualization; John J Ely, Mary Ann Raghanti, William D Hopkins, Patrick R Hof, Resources, Writing – review and editing; Chet C Sherwood, Conceptualization, Resources, Investigation, Writing – review and editing; Amy L Bauernfeind, Conceptualization, Investigation, Writing – original draft, Writing – review and editing; Courtney C Babbitt, Conceptualization, Formal analysis, Investigation, Writing – original draft, Writing – review and editing

Author ORCIDs

Katherine Rickelton http://orcid.org/0000-0001-6059-7505
Amy L Bauernfeind https://orcid.org/0000-0001-8518-3819
Courtney C Babbitt http://orcid.org/0000-0001-8793-4364

Decision letter and Author response

Decision letter https://doi.org/10.7554/eLife.70276.sa1
Author response https://doi.org/10.7554/eLife.70276.sa2

Additional files

Supplementary files

- Supplementary file 1. Species and sample information.
- Supplementary file 2. Percent identity thresholds used for orthology assignments for each species.

- Supplementary file 3. BUSCO Scores against Mammalian lineage (mammalia_odb10).
- Supplementary file 4. Eigenvalues, percent variance, and cumulative variance across disparate axes of the principal coordinates analysis (PCoA) of the 500 most variable genes from all data sampled.
- Supplementary file 5. Area occupied by each polygon representing either taxa or brain region in the principal coordinates analysis (PCoA) of the 500 most variable genes from all data sampled.
- Supplementary file 6. Enrichments are performed on an ordered query of differentially expressed (DE) genes where q<0.05 (this is not ranked on higher or lower expression, just absolute change).
- Supplementary file 7. Enrichments are performed on an ordered query of genes comparing those with a higher degree of fit for the Ornstein–Uhlenbeck (OU) model (stabilizing expression) with those that fit the Brownian Motion model (neural expression).
- Supplementary file 8. Species endocranial volumes (ECVs) were obtained from ***Isler et al., 2008*** and ***Coqueugniot and Hublin, 2012***. Reported volumes are averages of male and female data. ECVs are transformed to brain mass using the ***Stephan, 1960*** correction factor of 1.036 g/ml.
- Supplementary file 9. PC2 results from the brain size analysis by region.
- Supplementary file 10. Outlier Expression Analysis using the scoreGenes.R script in the EVEE-tools script set.
- MDAR checklist

Data availability

Sequencing data have been deposited in the Short Read Archive: BioProject PRJNA639850.

The following dataset was generated:

Author(s)	Year	Dataset title	Dataset URL	Database and Identifier
Rickelton K, Zintel TM, Pizzollo J, Miller E, Ely JJ, Raghanti MA, Hopkins WD, Hof PR, Sherwood CC, Bauernfeind AL, Babbitt CC	2023	Tempo and mode of gene expression evolution in the brain across primate tree	https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA639850	NCBI BioProject, PRJNA639850

The following previously published datasets were used:

Author(s)	Year	Dataset title	Dataset URL	Database and Identifier
Zhang Y, Sloan SA, Clarke LE, Caneda C	2015	RNA-Seq of human astrocytes	https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE73721	NCBI Gene Expression Omnibus, GSE73721
Consortium ENCODE Project	2016	single cell RNA-seq from pyramidal cell (ENCLB928LID)	https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE78331	NCBI Gene Expression Omnibus, GSE78331

References

Azevedo FAC, Carvalho LRB, Grinberg LT, Farfel JM, Ferretti REL, Leite REP, Jacob Filho W, Lent R, Herculano-Houzel S. 2009. Equal numbers of neuronal and nonneuronal cells make the human brain an isometrically scaled-up primate brain. *The Journal of Comparative Neurology* **513**:532–541. DOI: https://doi.org/10.1002/cne.21974, PMID: 19226510

Babbitt CC, Fedrigo O, Pfefferle AD, Boyle AP, Horvath JE, Furey TS, Wray GA. 2010. Both noncoding and protein-coding RNAs contribute to gene expression evolution in the primate brain. *Genome Biology and Evolution* **2**:67–79. DOI: https://doi.org/10.1093/gbe/evq002, PMID: 20333225

Bagot RC, Cates HM, Purushothaman I, Lorsch ZS, Walker DM, Wang J, Huang X, Schlüter OM, Maze I, Peña CJ, Heller EA, Issler O, Wang M, Song WM, Stein JL, Liu X, Doyle MA, Scobie KN, Sun HS, Neve RL, et al. 2016. Circuit-wide transcriptional profiling reveals brain region-specific gene networks regulating depression susceptibility. *Neuron* **90**:969–983. DOI: https://doi.org/10.1016/j.neuron.2016.04.015, PMID: 27181059

Barton RA. 2012. Embodied cognitive evolution and the cerebellum. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* **367**:2097–2107. DOI: https://doi.org/10.1098/rstb.2012.0112, PMID: 22734053

Blekhman R, Marioni JC, Zumbo P, Stephens M, Gilad Y. 2010. Sex-specific and lineage-specific alternative splicing in primates. *Genome Research* **20**:180–189. DOI: https://doi.org/10.1101/gr.099226.109, PMID: 20009012

Boddy AM, McGowen MR, Sherwood CC, Grossman LI, Goodman M, Wildman DE. 2012. Comparative analysis of encephalization in mammals reveals relaxed constraints on anthropoid primate and cetacean brain scaling. *Journal of Evolutionary Biology* **25**:981–994. DOI: https://doi.org/10.1111/j.1420-9101.2012.02491.x, PMID: 22435703

Brawand D, Soumillon M, Necsulea A, Julien P, Csárdi G, Harrigan P, Weier M, Liechti A, Aximu-Petri A, Kircher M, Albert FW, Zeller U, Khaitovich P, Grützner F, Bergmann S, Nielsen R, Pääbo S, Kaessmann H. 2011. The evolution of gene expression levels in mammalian organs. *Nature* **478**:343–348. DOI: https://doi.org/10.1038/nature10532, PMID: 22012392

Breschi A, Djebali S, Gillis J, Pervouchine DD, Dobin A, Davis CA, Gingeras TR, Guigó R. 2016. Gene-specific patterns of expression variation across organs and species. *Genome Biology* **17**:151. DOI: https://doi.org/10.1186/s13059-016-1008-y, PMID: 27391956

Chen J, Swofford R, Johnson J, Cummings BB, Rogel N, Lindblad-Toh K, Haerty W, Palma F di, Regev A. 2019. A quantitative framework for characterizing the evolutionary history of mammalian gene expression. *Genome Research* **29**:53–63. DOI: https://doi.org/10.1101/gr.237636.118, PMID: 30552105

Conway JR, Lex A, Gehlenborg N. 2017. UpSetR: an R package for the visualization of intersecting sets and their properties. *Bioinformatics* **33**:2938–2940. DOI: https://doi.org/10.1093/bioinformatics/btx364, PMID: 28645171

Coqueugniot H, Hublin JJ. 2012. Age-related changes of digital endocranial volume during human ontogeny: results from an osteological reference collection. *American Journal of Physical Anthropology* **147**:312–318. DOI: https://doi.org/10.1002/ajpa.21655, PMID: 22190338

Davis CA, Hitz BC, Sloan CA, Chan ET, Davidson JM, Gabdank I, Hilton JA, Jain K, Baymuradov UK, Narayanan AK, Onate KC, Graham K, Miyasato SR, Dreszer TR, Strattan JS, Jolanki O, Tanaka FY, Cherry JM. 2018. The encyclopedia of DNA elements (ENCODE): data portal update. *Nucleic Acids Research* **46**:D794–D801. DOI: https://doi.org/10.1093/nar/gkx1081, PMID: 29126249

DeCasien AR, Higham JP. 2019. Primate mosaic brain evolution reflects selection on sensory and cognitive specialization. *Nature Ecology & Evolution* **3**:1483–1493. DOI: https://doi.org/10.1038/s41559-019-0969-0, PMID: 31548648

Dos Santos SE, Medeiros M, Porfirio J, Tavares W, Pessôa L, Grinberg L, Leite REP, Ferretti-Rebustini REL, Suemoto CK, Filho WJ, Noctor SC, Sherwood CC, Kaas JH, Manger PR, Herculano-Houzel S. 2020. Similar microglial cell densities across brain structures and mammalian species: implications for brain tissue function. *The Journal of Neuroscience* **40**:4622–4643. DOI: https://doi.org/10.1523/JNEUROSCI.2339-19.2020, PMID: 32253358

Emera D, Yin J, Reilly SK, Gockley J, Noonan JP. 2016. Origin and evolution of developmental enhancers in the mammalian neocortex. *PNAS* **113**:E2617–E2626. DOI: https://doi.org/10.1073/pnas.1603718113, PMID: 27114548

ENCODE Project Consortium. 2012. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**:57–74. DOI: https://doi.org/10.1038/nature11247, PMID: 22955616

Fan C, Gao Y, Liang G, Huang L, Wang J, Yang X, Shi Y, Dräger UC, Zhong M, Gao T-M, Yang X. 2020. Transcriptomics of Gabra4 knockout mice reveals common NMDAR pathways underlying autism, memory, and epilepsy. *Molecular Autism* **11**:13. DOI: https://doi.org/10.1186/s13229-020-0318-9, PMID: 32033586

Fassio A, Patry L, Congia S, Onofri F, Piton A, Gauthier J, Pozzi D, Messa M, Defranchi E, Fadda M, Corradi A, Baldelli P, Lapointe L, St-Onge J, Meloche C, Mottron L, Valtorta F, Khoa Nguyen D, Rouleau GA, Benfenati F, et al. 2011. SYN1 loss-of-function mutations in autism and partial epilepsy cause impaired synaptic function. *Human Molecular Genetics* **20**:2297–2307. DOI: https://doi.org/10.1093/hmg/ddr122, PMID: 21441247

Funayama M, Nishioka K, Li Y, Hattori N. 2023. Molecular genetics of parkinson's disease: contributions and global trends. *Journal of Human Genetics* **68**:125–130. DOI: https://doi.org/10.1038/s10038-022-01058-5, PMID: 35821405

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z, Mauceli E, Hacohen N, Gnirke A, Rhind N, di Palma F, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N, et al. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology* **29**:644–652. DOI: https://doi.org/10.1038/nbt.1883, PMID: 21572440

Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D, Li B, Lieber M, MacManes MD, Ott M, Orvis J, Pochet N, Strozzi F, Weeks N, Westerman R, William T, Dewey CN, Henschel R, et al. 2013. De novo transcript sequence reconstruction from RNA-seq using the trinity platform for reference generation and analysis. *Nature Protocols* **8**:1494–1512. DOI: https://doi.org/10.1038/nprot.2013.084, PMID: 23845962

Hawrylycz MJ, Lein ES, Guillozet-Bongaarts AL, Shen EH, Ng L, Miller JA, van de Lagemaat LN, Smith KA, Ebbert A, Riley ZL, Abajian C, Beckmann CF, Bernard A, Bertagnolli D, Boe AF, Cartagena PM, Chakravarty MM, Chapin M, Chong J, Dalley RA, et al. 2012. An anatomically comprehensive atlas of the adult human brain transcriptome. *Nature* **489**:391–399. DOI: https://doi.org/10.1038/nature11405, PMID: 22996553

Hawrylycz M, Miller JA, Menon V, Feng D, Dolbeare T, Guillozet-Bongaarts AL, Jegga AG, Aronow BJ, Lee CK, Bernard A, Glasser MF, Dierker DL, Menche J, Szafer A, Collman F, Grange P, Berman KA, Mihalas S, Yao Z, Stewart L, et al. 2015. Canonical genetic signatures of the adult human brain. *Nature Neuroscience* **18**:1832–1844. DOI: https://doi.org/10.1038/nn.4171, PMID: 26571460

Henriques BJ, Katrine Jentoft Olsen R, Gomes CM, Bross P. 2021. Electron transfer flavoprotein and its role in mitochondrial energy metabolism in health and disease. *Gene* **776**:145407. DOI: https://doi.org/10.1016/j.gene.2021.145407, PMID: 33450351

Herculano-Houzel S. 2011. Scaling of brain metabolism with a fixed energy budget per neuron: implications for neuronal activity, plasticity and evolution. *PLOS ONE* **6**:e17514. DOI: https://doi.org/10.1371/journal.pone.0017514, PMID: 21390261

Hrvoj-Mihic B, Bienvenu T, Stefanacci L, Muotri AR, Semendeferi K. 2013. Evolution, development, and plasticity of the human brain: from molecules to bones. *Frontiers in Human Neuroscience* **7**:707. DOI: https://doi.org/10.3389/fnhum.2013.00707, PMID: 24194709

Huang DW, Sherman BT, Lempicki RA. 2009a. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Research* **37**:1–13. DOI: https://doi.org/10.1093/nar/gkn923, PMID: 19033363

Huang DW, Sherman BT, Lempicki RA. 2009b. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols* **4**:44–57. DOI: https://doi.org/10.1038/nprot.2008.211

Huerta-Cepas J, Dopazo J, Gabaldón T. 2010. ETE: a python environment for tree exploration. *BMC Bioinformatics* **11**:24. DOI: https://doi.org/10.1186/1471-2105-11-24, PMID: 20070885

Isler K, Christopher Kirk E, Miller JMA, Albrecht GA, Gelvin BR, Martin RD. 2008. Endocranial volumes of primate species: scaling analyses using a comprehensive and reliable data set. *Journal of Human Evolution* **55**:967–978. DOI: https://doi.org/10.1016/j.jhevol.2008.08.004, PMID: 18817943

Itō M. 2012. *The Cerebellum: Brain for an Implicit Self* Upper Saddle River, N.J: FT Press.

Kamilar JM, Cooper N. 2013. Phylogenetic signal in primate behaviour, ecology and life history. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* **368**:20120341. DOI: https://doi.org/10.1098/rstb.2012.0341, PMID: 23569289

Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler D. 2002. The human genome browser at UCSC. *Genome Research* **12**:996–1006. DOI: https://doi.org/10.1101/gr.229102, PMID: 12045153

Khaitovich P, Muetzel B, She X, Lachmann M, Hellmann I, Dietzsch J, Steigele S, Do HH, Weiss G, Enard W, Heissig F, Arendt T, Nieselt-Struwe K, Eichler EE, Pääbo S. 2004. Regional patterns of gene expression in human and chimpanzee brains. *Genome Research* **14**:1462–1473. DOI: https://doi.org/10.1101/gr.2538704, PMID: 15289471

Khaitovich P, Hellmann I, Enard W, Nowick K, Leinweber M, Franz H, Weiss G, Lachmann M, Pääbo S. 2005. Parallel patterns of evolution in the genomes and transcriptomes of humans and chimpanzees. *Science* **309**:1850–1854. DOI: https://doi.org/10.1126/science.1108296, PMID: 16141373

Khrameeva E, Kurochkin I, Han D, Guijarro P, Kanton S, Santel M, Qian Z, Rong S, Mazin P, Sabirov M, Bulat M, Efimova O, Tkachev A, Guo S, Sherwood CC, Camp JG, Pääbo S, Treutlein B, Khaitovich P. 2020. Single-cell-resolution transcriptome map of human, chimpanzee, bonobo, and macaque brains. *Genome Research* **30**:776–789. DOI: https://doi.org/10.1101/gr.256958.119, PMID: 32424074

King MC, Wilson AC. 1975. Evolution at two levels in humans and chimpanzees. *Science* **188**:107–116. DOI: https://doi.org/10.1126/science.1090005, PMID: 1090005

Konopka G, Friedrich T, Davis-Turak J, Winden K, Oldham MC, Gao F, Chen L, Wang GZ, Luo R, Preuss TM, Geschwind DH. 2012. Human-specific transcriptional networks in the brain. *Neuron* **75**:601–617. DOI: https://doi.org/10.1016/j.neuron.2012.05.034, PMID: 22920253

Konopka G, Roberts TF. 2016. Insights into the neural and genetic basis of vocal communication. *Cell* **164**:1269–1276. DOI: https://doi.org/10.1016/j.cell.2016.02.039, PMID: 26967292

Koropouli E, Kolodkin AL. 2014. Semaphorins and the dynamic regulation of synapse assembly, refinement, and function. *Current Opinion in Neurobiology* **27**:1–7. DOI: https://doi.org/10.1016/j.conb.2014.02.005, PMID: 24598309

Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology* **10**:R25. DOI: https://doi.org/10.1186/gb-2009-10-3-r25, PMID: 19261174

Li B, Dewey CN. 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**:323. DOI: https://doi.org/10.1186/1471-2105-12-323, PMID: 21816040

Ma S, Skarica M, Li Q, Xu C, Risgaard RD, Tebbenkamp ATN, Mato-Blanco X, Kovner R, Krsnik Ž, de Martin X, Luria V, Martí-Pérez X, Liang D, Karger A, Schmidt DK, Gomez-Sanchez Z, Qi C, Gobeske KT, Pochareddy S, Debnath A, et al. 2022. Molecular and cellular evolution of the primate dorsolateral prefrontal cortex. *Science* **377**:eabo7257. DOI: https://doi.org/10.1126/science.abo7257, PMID: 36007006

Martin RD. 1981. Relative brain size and basal metabolic rate in terrestrial vertebrates. *Nature* **293**:57–60. DOI: https://doi.org/10.1038/293057a0, PMID: 7266659

McCarthy DJ, Chen Y, Smyth GK. 2012. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Research* **40**:4288–4297. DOI: https://doi.org/10.1093/nar/gks042, PMID: 22287627

Miller IF, Barton RA, Nunn CL. 2019. Quantitative uniqueness of human brain evolution revealed through phylogenetic comparative analysis. *eLife* **8**:e41250. DOI: https://doi.org/10.7554/eLife.41250, PMID: 30702428

Montgomery SH, Mank JE. 2016. Inferring regulatory change from gene expression: the confounding effects of tissue scaling. *Molecular Ecology* **25**:5114–5128. DOI: https://doi.org/10.1111/mec.13824, PMID: 27564408
Munger EL, Edler MK, Hopkins WD, Hof PR, Sherwood CC, Raghanti MA. 2022. Comparative analysis of astrocytes in the prefrontal cortex of primates: insights into the evolution of human brain energetics. *The Journal of Comparative Neurology* **530**:3106–3125. DOI: https://doi.org/10.1002/cne.25387, PMID: 35859531
O'Brien RJ, Wong PC. 2011. Amyloid precursor protein processing and alzheimer's disease. *Annual Review of Neuroscience* **34**:185–204. DOI: https://doi.org/10.1146/annurev-neuro-061010-113613, PMID: 21456963
Perelman P, Johnson WE, Roos C, Seuánez HN, Horvath JE, Moreira MAM, Kessing B, Pontius J, Roelke M, Rumpler Y, Schneider MPC, Silva A, O'Brien SJ, Pecon-Slattery J. 2011. A molecular phylogeny of living primates. *PLOS Genetics* **7**:e1001342. DOI: https://doi.org/10.1371/journal.pgen.1001342, PMID: 21436896
Perry GH, Melsted P, Marioni JC, Wang Y, Bainer R, Pickrell JK, Michelini K, Zehr S, Yoder AD, Stephens M, Pritchard JK, Gilad Y. 2012. Comparative RNA sequencing reveals substantial genetic variation in endangered primates. *Genome Research* **22**:602–610. DOI: https://doi.org/10.1101/gr.130468.111, PMID: 22207615
Powell LE, Isler K, Barton RA. 2017. Re-evaluating the link between brain size and behavioural ecology in primates. *Proceedings. Biological Sciences* **284**:20171765. DOI: https://doi.org/10.1098/rspb.2017.1765, PMID: 29046380
Redies C, Hertel N, Hübner CA. 2012. Cadherins and neuropsychiatric disorders. *Brain Research* **1470**:130–144. DOI: https://doi.org/10.1016/j.brainres.2012.06.020, PMID: 22765916
Reimand J, Arak T, Adler P, Kolberg L, Reisberg S, Peterson H, Vilo J. 2016. g:Profiler-a web server for functional interpretation of gene lists (2016 update). *Nucleic Acids Research* **44**:W83–W89. DOI: https://doi.org/10.1093/nar/gkw199, PMID: 27098042
Rilling JK. 2014. Comparative primate neurobiology and the evolution of brain language systems. *Current Opinion in Neurobiology* **28**:10–14. DOI: https://doi.org/10.1016/j.conb.2014.04.002, PMID: 24835547
Robinson MD, McCarthy DJ, Smyth GK. 2010. edgeR: a bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**:139–140. DOI: https://doi.org/10.1093/bioinformatics/btp616, PMID: 19910308
Rohlfs RV, Harrigan P, Nielsen R. 2014. Modeling gene expression evolution with an extended ornstein-uhlenbeck process accounting for within-species variation. *Molecular Biology and Evolution* **31**:201–211. DOI: https://doi.org/10.1093/molbev/mst190, PMID: 24113538
Saitou N, Nei M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* **4**:406–425. DOI: https://doi.org/10.1093/oxfordjournals.molbev.a040454, PMID: 3447015
Seppey M, Manni M, Zdobnov EM. 2019. Busco: Assessing genome assembly and annotation completeness. *Methods in Molecular Biology* **1962**:227–245. DOI: https://doi.org/10.1007/978-1-4939-9173-0_14, PMID: 31020564
Sherwood CC, Miller SB, Karl M, Stimpson CD, Phillips KA, Jacobs B, Hof PR, Raghanti MA, Smaers JB. 2020. Invariant synapse density and neuronal connectivity scaling in primate neocortical evolution. *Cerebral Cortex* **30**:5604–5615. DOI: https://doi.org/10.1093/cercor/bhaa149, PMID: 32488266
Shil SK, Kagawa Y, Umaru BA, Nanto-Hara F, Miyazaki H, Yamamoto Y, Kobayashi S, Suzuki C, Abe T, Owada Y. 2021. Ndufs4 ablation decreases synaptophysin expression in hippocampus. *Scientific Reports* **11**:10969. DOI: https://doi.org/10.1038/s41598-021-90127-4, PMID: 34040028
Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**:3210–3212. DOI: https://doi.org/10.1093/bioinformatics/btv351, PMID: 26059717
Slater PG, Ramirez VT, Gonzalez-Billault C, Varela-Nallar L, Inestrosa NC. 2013. Frizzled-5 receptor is involved in neuronal polarity and morphogenesis of hippocampal neurons. *PLOS ONE* **8**:e78892. DOI: https://doi.org/10.1371/journal.pone.0078892, PMID: 24205342
Smaers JB, Turner AH, Gómez-Robles A, Sherwood CC. 2018. A cerebellar substrate for cognition evolved multiple times independently in mammals. *eLife* **7**:e35696. DOI: https://doi.org/10.7554/eLife.35696, PMID: 29809137
Smaers JB, Rothman RS, Hudson DR, Balanoff AM, Beatty B, Dechmann DKN, de Vries D, Dunn JC, Fleagle JG, Gilbert CC, Goswami A, Iwaniuk AN, Jungers WL, Kerney M, Ksepka DT, Manger PR, Mongle CS, Rohlf FJ, Smith NA, Soligo C, et al. 2021. The evolution of mammalian brain size. *Science Advances* **7**:eabe2101. DOI: https://doi.org/10.1126/sciadv.abe2101, PMID: 33910907
Somel M, Franz H, Yan Z, Lorenc A, Guo S, Giger T, Kelso J, Nickel B, Dannemann M, Bahn S, Webster MJ, Weickert CS, Lachmann M, Pääbo S, Khaitovich P. 2009. Transcriptional neoteny in the human brain. *PNAS* **106**:5743–5748. DOI: https://doi.org/10.1073/pnas.0900544106, PMID: 19307592
Sousa AMM, Meyer KA, Santpere G, Gulden FO, Sestan N. 2017a. Evolution of the human nervous system function. *Structure, and Development. Cell* **170**:226–247. DOI: https://doi.org/10.1016/j.cell.2017.06.036
Sousa AMM, Zhu Y, Raghanti MA, Kitchen RR, Onorati M, Tebbenkamp ATN, Stutz B, Meyer KA, Li M, Kawasawa YI, Liu F, Perez RG, Mele M, Carvalho T, Skarica M, Gulden FO, Pletikos M, Shibata A, Stephenson AR, Edler MK, et al. 2017b. Molecular and cellular reorganization of neural circuits in the human lineage. *Science* **358**:1027–1032. DOI: https://doi.org/10.1126/science.aan3456, PMID: 29170230
Stephan H. 1960. Methodische studien uber den quantitativen vergleicharchitektonischer struktureinheiten des gehirns. *Zeitschrift Fur Wissenschartliche Zoologie* **164**:143–172.
Stout D, Hecht EE. 2017. Evolutionary neuroscience of cumulative culture. *PNAS* **114**:7861–7868. DOI: https://doi.org/10.1073/pnas.1620738114, PMID: 28739892

Sudmant PH, Alexis MS, Burge CB. 2015. Meta-analysis of RNA-seq expression data across species, tissues and studies. *Genome Biology* **16**:287. DOI: https://doi.org/10.1186/s13059-015-0853-4, PMID: 26694591

Szeri F, Lundkvist S, Donnelly S, Engelke UFH, Rhee K, Williams CJ, Sundberg JP, Wevers RA, Tomlinson RE, Jansen RS, van de Wetering K. 2020. The membrane protein ANKH is crucial for bone mechanical performance by mediating cellular export of citrate and ATP. *PLOS Genetics* **16**:e1008884. DOI: https://doi.org/10.1371/journal.pgen.1008884, PMID: 32639996

Uhlén M, Fagerberg L, Hallström BM, Lindskog C, Oksvold P, Mardinoglu A, Sivertsson Å, Kampf C, Sjöstedt E, Asplund A, Olsson I, Edlund K, Lundberg E, Navani S, Szigyarto CAK, Odeberg J, Djureinovic D, Takanen JO, Hober S, Alm T, et al. 2015. Proteomics: tissue-based map of the human proteome. *Science* **347**:1260419. DOI: https://doi.org/10.1126/science.1260419, PMID: 25613900

Yapar E, Saglican E, Dönertaş HM, Özkurt E, Yan Z, Hu H, Guo S, Erdem B, Rohlfs RV, Khaitovich P, Somel M. 2021. Convergent Evolution of Primate Testis Transcriptomes Reflects Mating Strategy. *bioRxiv*. DOI: https://doi.org/10.1101/010553

Zhang Y, Sloan SA, Clarke LE, Caneda C, Plaza CA, Blumenthal PD, Vogel H, Steinberg GK, Edwards MSB, Li G, Duncan JA, Cheshier SH, Shuer LM, Chang EF, Grant GA, Gephart MGH, Barres BA. 2016. Purification and characterization of progenitor and mature human astrocytes reveals transcriptional and functional differences with mouse. *Neuron* **89**:37–53. DOI: https://doi.org/10.1016/j.neuron.2015.11.013, PMID: 26687838