Systematic detection of divergent brain proteins in human evolution and their roles in cognition

Guillaume Dumas[1,*], Simon Malesys[1] and Thomas Bourgeron[1]

Affiliations:

1 Human Genetics and Cognitive Functions, Institut Pasteur, UMR3571 CNRS, Université de Paris, Paris, (75015) France

* Corresponding author: guillaume.dumas@pasteur.fr

Abstract

The human brain differs from that of other primates, but the underlying genetic mechanisms remain unclear. Here we measured the evolutionary pressures acting on all human protein-coding genes (N=17,808) based on their divergence from early hominins such as Neanderthal, and non-human primates. We confirm that genes encoding brain-related proteins are among the most conserved of the human proteome. Conversely, several of the most divergent proteins in humans compared to other primates are associated with brain-associated diseases such as micro/macrocephaly, dyslexia, and autism. We identified specific expression profiles of a set of divergent genes in ciliated cells of the cerebellum, that might have contributed to the emergence of fine motor skills and social cognition in humans. This resource is available at http://neanderthal.pasteur.fr and can be used to estimate evolutionary constraints acting on a set of genes and to explore their relative contribution to human traits.

Introduction

Homo sapiens possess high abilities to perform complex cognitive tasks and to communicate with peers[1]. Anatomical brain differences between humans and other primates are well documented (e.g. cortex size, prefrontal white matter thickness, lateralization), but the mode in which the human brain has evolved remains controversial[2]. For example, modern humans have large and globular brains that distinguish them from their extinct *Homo* relatives, but it remains unknown when and how brain globularity evolved and how it relates to evolutionary brain size increase. A recent study analyzing endocranial casts of *Homo sapiens* fossils indicate that 300,000 years ago, brain size in early *Homo sapiens* already fell within the range of present-day humans[3]. Brain shape, however, evolved gradually within the *Homo sapiens* lineage, reaching present-day human variation between about 100,000 and 35,000 years ago. It has also been suggested that the enlargement of the prefrontal cortex relative to motor cortex in humans was mirrored in the cerebellum with an enlargement of the regions of the cerebellum connected to the prefrontal cortex[4]. These anatomical processes of tandem evolution in the brain paralleled the emergence of motor and cognitive abilities such as bipedalism, planning, language, as well as social awareness that are especially well developed in humans.

Genetic diversity between primates has surely contributed to these brain and cognitive differences, but the contributing genes or variants remain largely unknown. Indeed, demonstrating that a genetic variant is adaptive requires strong evidence both at the genetic and functional levels. There are few human specific genes such as *SRGAP2C*[5], *ARHGAP11B*[6] or *NOTCH2NL*[7] that recently appeared through gene duplication in the *Homo* lineage[8] and remarkably their expression in the mouse brain increases cortical development[6,7,9,10]. In

addition, several genes involved in brain function display accelerated evolution of their coding regions in humans, for example *FOXP2 has been* associated with verbal apraxia and *ASPM* with microcephaly[11,12]. Functional studies also showed that mice carrying a "humanized" version of FOX2 display qualitative changes in ultrasonic vocalization[13]. However, these reports targeting only specific genes sometimes provide contradictory results[14] and other studies have reported a greater sequence conservation of the brain protein-coding genes compared to other tissues[15–17], suggesting that the main substrate of evolution in the brain reside in regulatory changes of gene expression[18–20] and splicing[21]. It was also shown that more genes underwent positive selection in chimpanzee evolution than in human evolution contradicting the widespread brain-gene acceleration hypothesis of human cognition origins[22]. Several studies have recently explored the most constraint genes during primate evolution or in human populations to better estimate the pathogenicity of variants identified in patients with genetic disorders[23,24]. In contrast, few studies made a systematic detection of the divergent genes during primate evolution [25,26]. Here, we provide an exhaustive screening of all protein-coding genes for conservation and divergence from our common primate ancestor, and take advantage of rich datasets of brain single-cell transcriptomics, proteomics and imaging to investigate their relations to brain structure, function, and diseases.

Results

High conservation of brain protein-coding genes

We first compared the sequence of modern, archaic humans, and other primates to their common ancestor, and extract a measure of evolution for 17,808 protein-coding genes in the modern human genome (Fig. 1a, see also Supplementary Fig. 1 and 2)[27]. This resource is

available online at http://neanderthal.pasteur.fr. Our measure is derived from one of the most popular and reliable measures of evolutionary pressures on protein-coding regions, the dN/dS ratio[28] —also called ω. This measure compares the rates of non-synonymous and synonymous mutations of coding sequences. If there are more non-synonymous mutations than expected, there is divergence, if less, there is conservation. First, we estimate the dN and dS for all the genes that are 1:1 orthologous since the evolutionary constraints are relaxed in duplicated genes[29] (note: only the Y chromosome was ignored from the analyses). Then, we adjusted the dN/dS ratio for biases induced by variations of mutations rates with the GC content of codons. Finally, we renormalized across the genome the values obtained for each taxon. The final ω_{GC12} is a form of Z-score corrected for GC content that quantifies the unbiased relative divergence of genes compared to the ancestral primate genome.

Using the ω_{GC12} on all protein coding genes for each hominin —*Homo sapiens*, Denisovan, Neanderthal, and *Pan troglodytes*— we identified two distinct clusters (Fig. 1b and Supplementary Table 1): one with divergent proteins, enriched in olfactory genes (OR=1.48, p=8.4e-9), and one with conserved proteins, enriched in brain related biological functions (Fig. 1c and Supplementary Table 2). This second cluster reveals especially a strong conservation of proteins implicated in nervous system development (OR=1.2, p=2.4e-9) and synaptic transmission (OR=1.35, p=1.7e-8).

We then focused on the genes expressed in different body tissues by combining different RNAseq (Illumina Bodymap2 and GTEx), microarray and proteomics datasets. For expression data, we characterized the specificity of genes by normalizing their profile across tissues (Supplementary Fig. 3). The results confirmed with three different datasets that protein-coding genes expressed in the brain were more conserved (Wilcoxon rc=-0.1, p=4.1e-

12) than genes related to other parts of the body, testis being the most divergent (Wilcoxon rc=0.3, p=7.8e-11; Fig. 1d, see also Supplementary Fig. 4 and 5).

Conservation of proteins related to nervous system substructure and neuronal functions

Within the central nervous system, three substructures appeared to evolve under the strongest degree of purifying selection at the protein sequence level, i.e. highly conserved: the cerebellum (Wilcoxon rc=-0.29, p=5.5e-6), the amygdala (Wilcoxon rc=-0.11, p=4.1e-6), and more surprisingly the prefrontal cortex (Wilcoxon rc=-0.1, p=5.7e-10; Fig. 2a, see also Supplementary Table 3). Indeed, the prefrontal cortex is proposed as one of the most divergent brain structure in human evolution[30] and is associated with high cognitive functions[31]. Only one brain structure was more divergent than expected: the superior cervical ganglion (Wilcoxon rc=0.22, p=1e-6), which provides sympathetic innervation to many organs and is associated with the archaic functions of fight or flight response. Among the divergent genes expressed in the superior cervical ganglion, CARF is specifically divergent in the genus *Homo*. Interestingly, this calcium responsive transcription factor regulates neuronal activity-dependent expression of BDNF[32] as well as a set of singing-induced genes in the song nuclei of the zebra finch, a vocal learning songbird[33]. The gene has a dN/dS of 2.44 and is among the most divergent proteins expressed in the human brain.

Regarding the brain gene ontologies, all pathways are overall highly conserved, but genes involved in glutamatergic and GABAergic neurotransmissions are on average more conserved (Wilcoxon rc=-0.25; p=9.8e-6) than those participating to dopamine and peptide neurotransmission and to intracellular trafficking (Fig. 2b, see also Supplementary Fig. 6 and Supplementary Table 3). Integrating the recently released ontology of the synapse provided

by the SynGO consortium (http://syngoportal.org), not only we confirmed this global strong conservation of the synapse, but also revealed how it especially concerns the trans-synaptic signaling processes (Wilcoxon rc=-0.21, p=4.5e-5) as well as postsynaptic (rc=-0.56, p=6.3e-8) and presynaptic membranes (Wilcoxon: rc=-0.56, p=7e-8 ; Fig. 2c,d).

The divergent proteins and their correlation with brain expression and functional activity

In order to focus on the most divergent proteins, we took the proteins situated at the extremes of the ω_{GC12} distribution (>2SD; Fig. 3a; Supplementary Table 4) and which have been fixed in the modern *Homo sapiens* population (Neutrality Index<1). Among those 352 highly divergent proteins, 126 were related to the brain (Fisher exact OR=0.66, p=1e-4), listed as a synaptic proteins[34,35], specifically expressed in the brain (+2SD of specific expression) or related to a brain disease (listed in OMIM: https://www.omim.org or HPO: https://hpo.jax.org/app/). For comparison, we also extracted the 427 most conserved proteins, among which 290 were related to the brain categories listed above (enrichment for brain genes, Fisher exact OR=1.26, p=0.0032).

Using these 352 highly conserved and 427 highly divergent genes, we first confirmed using the Brainspan data available on the Specific expression Analysis (SEA) that genes expressed in the cerebellum and the cortex were enriched in conserved genes (Supplementary Figure 7). Despite this conservation on average, we nevertheless identified several divergent genes, some even specifically divergent in *Homo sapiens* such as PBXIP1 (dN/dS=2.35) which has a role in cell migration and proliferation, but is also an astroglial progenitor cell marker during human embryonic development[36]. We investigated the pattern of expression of the highly divergent genes through the lens of the Allen Brain atlas[37]. We

compared among the six donors how reproducible was the difference of expression between highly divergent and conserved genes and found a cluster of brain regions (hypothalamus, cerebral nuclei, and cerebellum) expressing more specifically highly divergent genes (Fig 4). Among the divergent genes, the most specifically expressed in those subcortical structures were two members of the NACA family (NACAD; dN/dS =1.31 and NACA; dN/dS =2.92), which play a role in protein transport and targeting, as well as KIAA1191 (dN/dS =1.31) and KIAA0319 (dN/dS =3.89), both involved in the regulation of neuronal survival, differentiation, and axonal outgrowth[38,39]. Using single-cell transcriptomics[40,41], we finally found that ciliated cells in the cerebellum and choroid cells in the cortex were the two main cell types displaying specific divergence (Fig 5a; Supplementary Figure 8). Combining our analyses across *Homo sapiens*, Neanderthal, Denisovan, and *Pan troglodytes*, we also found among proteins specifically expressed in those ciliated cells some which were specifically divergent in modern humans (Fig. 5b): C11orf16, EFCAB12, LCA5, and NSUN7. Interestingly, LCA5 (dN/dS=1.6) is directly involved in centrosomal or ciliary function, and has been shown to interact with genes implicated in Joubert Syndrome[42]. Some proteins linked to brain morphology disorders also displayed a divergence specifically in the hominin lineage: TTLL6, DYNC2LI1, and TOP3A respectively implicated in Joubert syndrome, ciliopathy, and microcephaly. TOP3A is particularly interesting since it is strongly conserved in *Pan troglodytes* and two hominin specific variants (RNA NM_004618, R831Q and E959G) are located in the vicinity of pathogenic mutations causing microcephaly [43].

Finally, to test potential association with brain functions, we extracted from NeuroVault[44] 19,244 brain imaging results from fMRI-BOLD studies (T and Z score maps) and examined how their spatial patterns correlate with those of genes expression in the Allen Brain atlas[45]. The correlation between brain activity and divergent genes expression was

stronger in subcortical structures than in the cortex (Wilcoxon rc=0.14, p=2.5e-248). The brain activity maps that correlate to the expression pattern of the divergent genes (see Supplementary Table 6 for details) were enriched in social tasks (empathy, emotion recognition, theory of mind, language; Fisher exact p=2.9e-20, odd ratio=1.72; see Supplementary Figure 9 for illustration).

The divergent proteins and their connection with brain disorders

Our systematic analysis revealed that highly constrained proteins were more associated with brain diseases or traits compared to divergent proteins especially with intellectual disability and autism but also diseases associated with myelin or encephalopathy (Figure 3b). Those highly conserved proteins associated with brain diseases include tubulins (TUBA1A, TUBB3, TUBB4A), dynamin (DNM1), chromatin remodeling proteins (SMARCA4) and signaling molecules such as AKT1, DVL1, NOTCH1 and its ligand DLL1, all associated with different forms of neurodevelopmental disorders (Supplementary Table 4). We also found 31 highly divergent proteins that are associated with several human diseases or conditions such as micro/macrocephaly, autism or dyslexia.

When we compared humans and chimpanzees with our common primate ancestor, several proteins associated to micro/macrocephaly displayed different patterns of evolution in humans and chimpanzees (Fig. 6). Some proteins displayed a divergence specifically in the hominin lineage (AHI1, ASXL1, CSPP1, DAG1, FAM111A, GRIP1, NHEJ1, QDPR, RNF135, RNF168, SLX4, TCTN1, and TMEM70) or alternatively in the chimpanzee (ARHGAP31, ATRIP, CPT2, CTC1, HDAC6, HEXB, KIF2A, MKKS, MRPS22, RFT1, TBX6, and WWOX). The protein phosphatase PPP1R15B associated with microcephaly was divergent from the common

primate ancestor in both taxa. None of the proteins related to micro/macrocephaly appeared divergent only in *Homo sapiens* (Fig. 6).

We also found divergent proteins associated with communication disorders (Fig. 3c), such as autism (CNTNAP4, AHI1, FAN1, SNTG2 and GRIP1) and dyslexia (KIAA0319). Interestingly, these proteins are divergent from the common primate ancestor only in the hominin lineage and highly conserved in all other taxa (Fig. 7). They all have roles related to neuronal connectivity (neuronal migration and synaptogenesis) and were within the human brain relatively more expressed in the cerebellum except GRIP1 that is almost only restricted to the cortex.

Among the genes associated with autism, CNTNAP4 is a member of the neurexin protein family that plays a role in proper neurotransmission in the dopaminergic and GABAergic systems[46]. SNTG2 is a cytoplasmic peripheral membrane protein that binds to NLGN3 and NLGN4X, two proteins associated with autism[47] and several copy-number variants affecting *SNTG2* were identified in patients with autism[48]. GRIP1 (Glutamate receptor interacting protein 1) is also associated with microcephaly and codes a synaptic scaffolding protein that interacts with glutamate receptors. Variants of this gene were recurrently associated with autism[49].

Interestingly, we also detected that the dyslexia-susceptibility protein KIAA0319, involved in axon growth inhibition[38,50], is among the most divergent brain proteins in humans compared to our common primate ancestor (dN/dS=3.9). The role of KIAA0319 in dyslexia is still debated, but its rapid evolution in the hominoid lineage warrants further genetic and functional studies.

Finally, several genes display very high divergence in *Homo sapiens*, but their functions or their association to diseases remain unknown. For example, the zinc finger protein ZNF491 (dN/dS=4.7) is highly expressed in the cerebellum and displays structural similarities with a chromatin remodeling factor, but its biological role remains to be characterized. Another example is CCP110, a centrosomal protein such as ASPM, but not associated to a disease. Based on its function, this divergent protein CCP110 could be a compelling candidate for microcephaly in humans. A complete list of the most conserved and divergent proteins is available in Supplementary Table 4 and on the companion website.

Systematic analysis of Neanderthal introgression.

Finally, we also looked at the recent evolutionary event of Neanderthal introgression. Combining the dataset of Sankararaman[51] with our systematic screening of coding sequences, we derived for 16,020 genes the level of Neanderthal introgression (also available on the website). We conduced similar analyses as with ω_{GC12} (see Supplementary Table 7) and despite no correlation between ω_{GC12} and introgression, we obtained opposite trends at the tissue level: genes expressed in the testes appeared as the less introgressed tissues of the body (Wilcoxon rc=-0.09, p=1.5e-10), while genes associated to the adult brain in RNAseq (Wilcoxon rc=0.16, p=1.2e-4) and Mass Spectrometry (Wilcoxon rc=0.26, p=2.6e-6) appeared more introgressed than expected (Supplementary Figure 10). At the nervous system level, the spinal cord (Wilcoxon rc=0.3, p=1.9e-5) and the retina (Wilcoxon rc=0.15, p=3e-9) revealed strong introgression, while the pineal gland had less introgression than expected (Wilcoxon, pineal during day: rc=-0.14, p=6.2e-15; pineal during night: rc=-0.12, p=6.5e-10; Supplementary Figure 11). At the functional level, the proteins associated with the

glutamatergic (Wilcoxon rc=0.38, p=3e-5) and the serotonin (Wilcoxon rc=0.29, p=1.8e-3) pathways were also highly introgressed (Supplementary Figure 12).

Discussion

Divergent proteins and brain sizes in primates

Several proteins were previously proposed to play a major role in the expansion of the size of the human brain. Some of these proteins such as ARHGAP11B, SRGAP2C and NOTCH2NL[7] were found specific to humans since their genes were recently duplicated[52]. Other studies proposed that high divergence in genes involved in micro/macrocephaly could have contributed to the dramatic change in brain size during primate evolution[25,53]. Several of these genes apparently displayed an accelerated evolution in humans such as the ASPM[54] and MCPH1[55]. However, the adaptive evolution of these genes was previously questioned[56] and in our analysis neither of these two proteins are on the list of the highly divergent proteins (their dN/dS value are indeed below 0.8).

Conversely, our systematic detection could identify the most divergent proteins in humans for micro/macrocephaly and the top 10 proteins were FAM111A, AHI1, CSPP1, TCTN1, DAG1, TMEM70, ASXL1, RNF168, NHEJ1, GRIP1. This list of divergent proteins associated with micro/macrocephaly in humans can be used to select the best candidate human specific gene/variants for further genetic and functional analyses in order to better estimate their contribution to the emergence of anatomical difference between humans and other primates.

Some of these proteins could have contributed to brain size and to other morphological features such as difference in skeleton development. For example, the divergent proteins FAM111A (dN/dS=2.99) and ASXL1 (dN/dS=1.83) are associated respectively with macrocephaly and microcephaly. Patients with dominant mutation of FAM111A are diagnosed with Kenny-Caffey syndrome (KCS). They display impaired skeletal development with small and dense bones, short stature, primary hypoparathyroidism with hypocalcemia and a prominent forehead[57]. While the function of FAM111A remains largely unknown, this protein appears to be crucial to a pathway that governs parathyroid hormone production, calcium homeostasis, and skeletal development and growth. On the other hand, patients with dominant mutations of *ASXL1* are diagnosed with Bohring-Opitz syndrome, a malformation syndrome characterized by severe intrauterine growth retardation, intellectual disability, trigonocephaly, hirsutism, and flexion of the elbows and wrists with deviation of the wrists and metacarpophalangeal joints[58]. ASXL1 encodes a chromatin protein required to maintain both activation and silencing of homeotic genes.

Remarkably, three proteins AHI1, CSPP1 and TCTN1 are in the top 5 with dN/dS>2 and are required for both cortical and cerebellar development in humans and associated with Joubert syndrome. Children diagnosed with this recessive disease present with agenesis of the cerebellar vermis and difficulties coordinating movements. AHI1 is a positive modulator of classical WNT/Ciliary signaling. CSPP1 is involved in the cell-cycle-dependent microtubule organization and TCTN1 is a regulator of Hedgehog during development.

AHI1 was previously reported as a gene under positive evolutionary selection along the human lineage[59,60], but to our knowledge, neither CSPP1 and TCTN1 were previously detected as divergent proteins during primate evolution. Interestingly, it was proposed that

the accelerated evolution of AHI1 required for ciliogenesis and axonal growth could have played a role in the development of unique motor capabilities in humans, such as bipedalism[53]. Our findings provide further support for accelerated evolution of a set of genes associated with ciliogenesis. Indeed, we found that three additional proteins CSPP1, TTLL6, and TCTN1 involved in Joubert syndrome are among the most divergent proteins during human evolution and our single cell analysis revealed that ciliated cells are the main category of cerebellar cells expressing divergent genes.

The possible link between a change in the cerebellum genetic makeup and the evolution of human cognition

Although the emergence of a large brain was probably an important step for human cognition, other parts of the brain such as the cerebellum might have also played an important role in both motricity and cognition. In this study, we showed that proteins expressed in the cerebellum were among the most conserved proteins in humans. Nevertheless, we also found a set of divergent proteins with relatively high expression in the cerebellum and/or affecting cerebellar function when mutated. As discussed above, several genes associated with Joubert syndrome such as AHI1, CSPP1, TTLL6, and TCTN1 display protein divergence in humans and are important for cerebellar development. In addition, among the most divergent proteins expressed in the brain, CNTNAP4, FAN1, SNTG2, and KIAA0319 also display high expression in the cerebellum and have been associated with communication disorders such as autism and dyslexia.

Within the cerebellum, the lateral zone, also known as neocerebellum, is novel to mammals and has expended during primate evolution[61,62]. Specifically, it was shown that this

part of the cerebellum connected to the prefrontal cortex was enlarged, relative to the cerebellar lobules connected to the motor cortex[4]. Interestingly, the largest difference in the endocranial shapes of modern humans and Neanderthals are found in the cerebellum[63], although other regions are also relatively larger in modern humans than in Neanderthals, including parts of the prefrontal cortex and the occipital and temporal lobes[64].

In this line, a recent study detected Neanderthal alleles nearby two genes, UBR4 on chromosome 1 and PHLPP1 on chromosome 8, which could be linked to difference in endocranial globularity[64]. These genes are involved in neurogenesis and myelination, respectively, and UBR4 is most specifically expressed in the cerebellum. In our analysis, the coding regions of these genes are highly conserved (dN/dS<0.13) suggesting that regulatory variants rather than coding sequence variants might be involved.

In humans, the cerebellum is associated with higher cognitive functions such as visual-spatial skills, planning of complex movements, procedural learning, attention switching, and sensory discrimination[65]. It has a key role for temporal processing[66] and for being able to anticipate and control behavior with both implicit and explicit mechanisms[65]. It is therefore expected that a change in the genetic makeup of the cerebellum might have been of great advantage for the emergence of specific human cognition.

Despite this possible link between the cerebellum and the emergence of human cognition, this part of the brain was neglected compared to the cortex where most of the functional studies investigating the role of human specific genes/variants have been conducted. For example, while in humans the expression of SRGAP2C is almost restricted to the cerebellum, its ectopic expression was studied in the mouse cortex[5,10] where it triggers human-like neuronal characteristics, such as increased dendritic spine length and density.

Hence, we propose that an exploration of specific human genes/variants to the development and the function of the cerebellum might shed new light on the evolution of human cognition.

Perspectives

Our systematic analysis of protein sequence diversity confirms that proteins related to brain function are among the most conserved proteins of the human proteome. The set of divergent proteins that we identified could have played specific roles in the evolution of human cognition by modulating brain size, neuronal migration and/or synaptic physiology, but further genetic and functional studies should shed new light on the role of these divergent proteins. Beyond the brain, this resource will be also helpful to estimate the level of evolutionary pressure acting on genes related to other biological pathways, especially those displaying signs of positive selections during primate evolution such as the reproductive and immune systems.

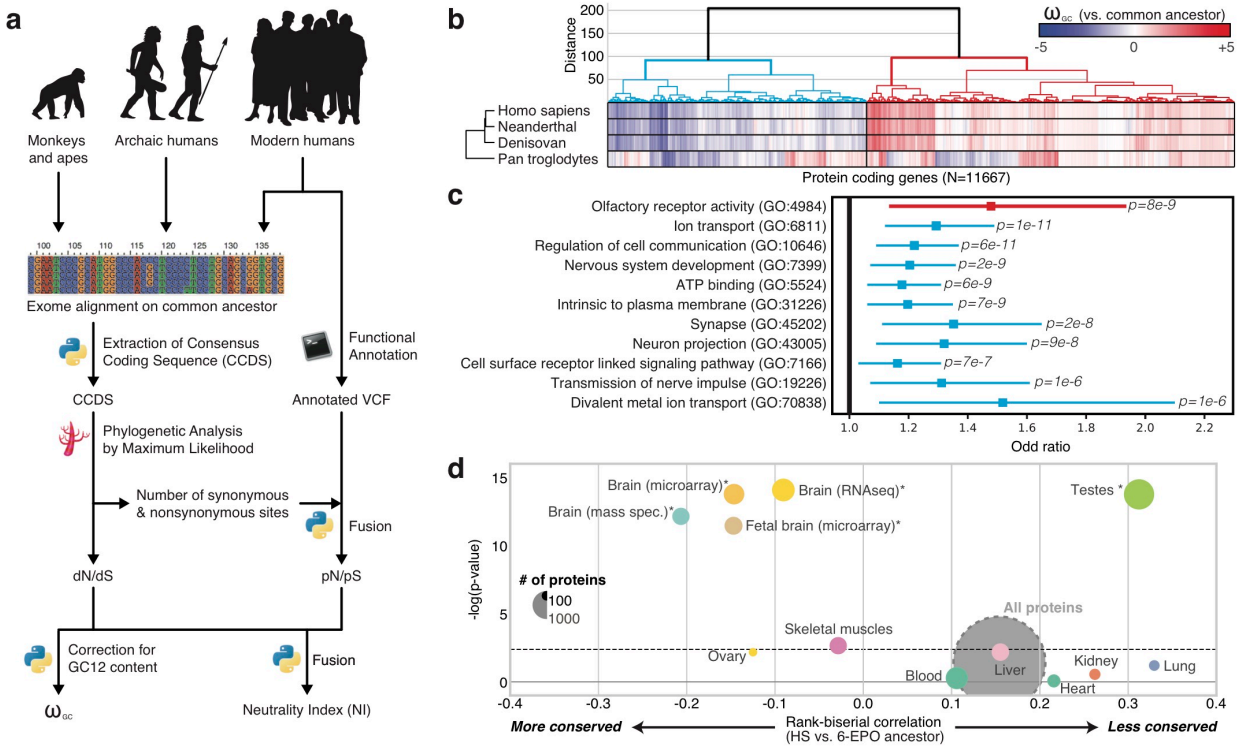

Figure 1 Evolution of protein-coding genes across tissues and biological functions. (**a**) Analyses pipeline for extraction of ω_{GC12}, a corrected and normalized measure of evolution of the protein coding genes that behaves as a Z-score that takes into account the GC content of codons. (**b**) Hierarchical clustering on ω_{GC12} across all protein coding genes (1:1 orthologs in hominins with medium coverage; See Supplementary Table 1). (**c**) Gene Ontology (GO) enrichments for the red and the blue clusters in panel b (See Supplementary Table 2 for all GO). (**d**) Funnel plot summarizing the evolution of protein-coding genes specifically expressed in different tissues of the human body (Supplementary Table 3). Horizontal dash line stands for the Bonferroni threshold. Stars indicate set of genes passing statistical significance for multiple comparison correction with a bootstrap that takes into account GC12 content and length of the coding sequence. HS: *Homo sapiens*; 6-EPO ancestor: the reconstructed ancestral genome of primates based on alignments of human, chimpanzee, gorilla, orangutan, rhesus macaque, and marmoset genomes.

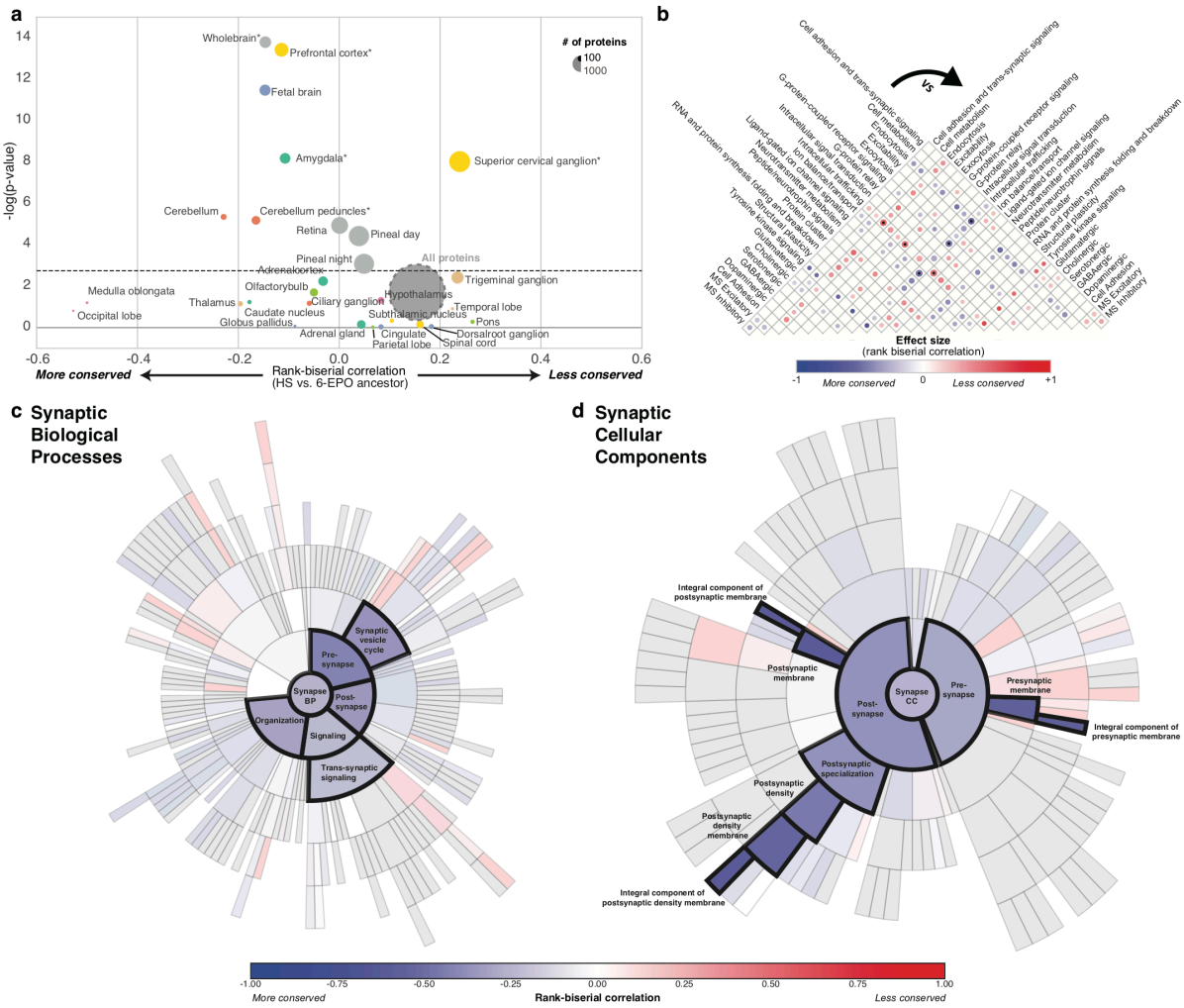

Figure 2 Evolution of protein-coding genes related to the brain. (**a**) Funnel plot summarizing the evolution of protein-coding genes specifically expressed in brain substructures; horizontal dash line stands for the Bonferroni threshold. Stars indicate set of genes passing statistical significance for multiple comparison correction with a bootstrap; (**b**) Matrix summarizing the effect size of the difference of protein-coding gene divergence between synaptic functions; color cells indicate Mann-Whitney comparison with nominal p-value <0.05. Black dots indicate comparison passing statistical significance for the Bonferroni threshold. (**c, d**) SynGO sunburst plots showing nested statistically conserved (blue) biological processes and cellular components of the synapse.

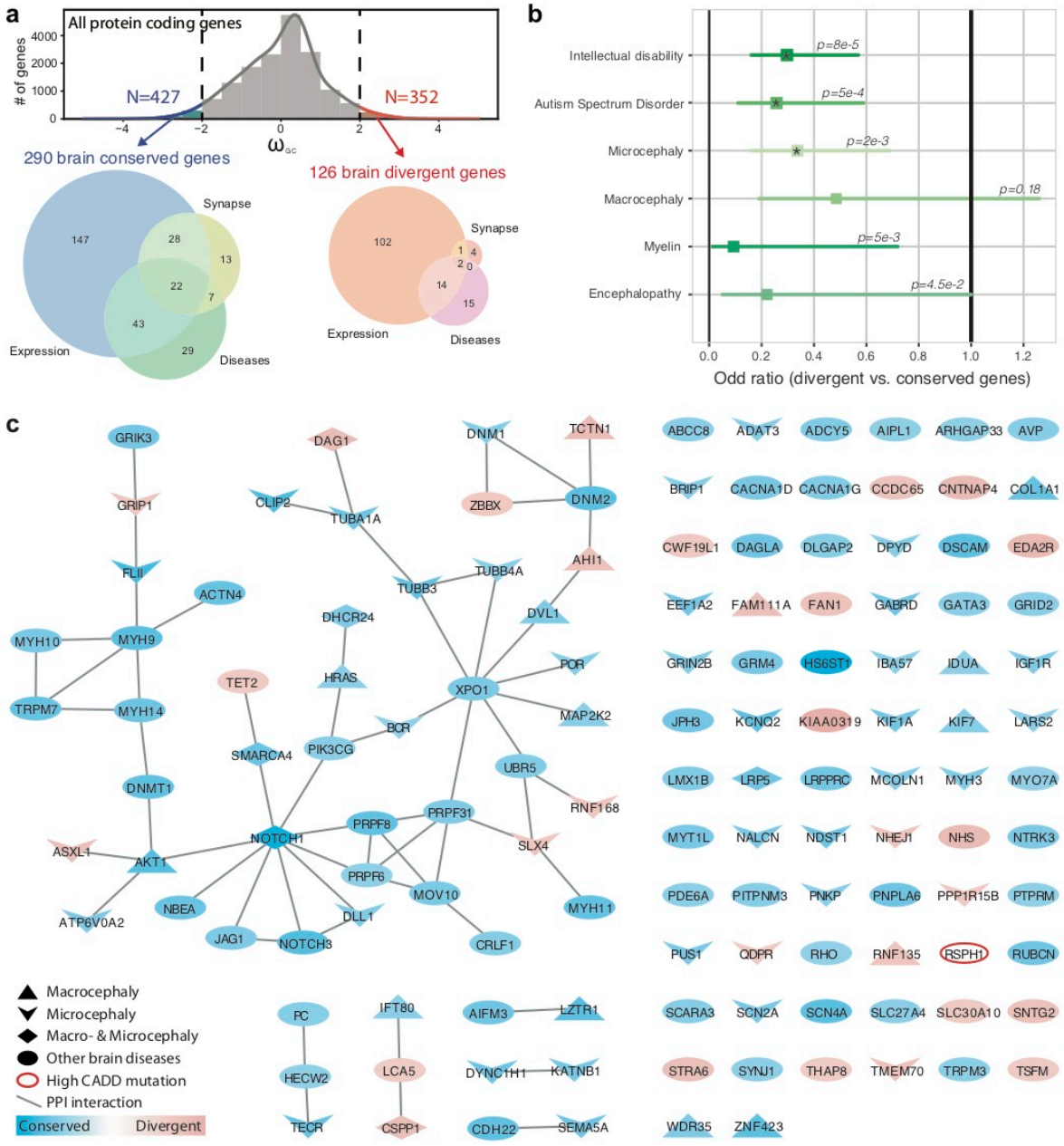

Figure 3 Brain proteins and human diseases. (**a**) Distribution of ω_{GC12} and Venn diagrams describing the most conserved and divergent proteins with specific brain expression, related to the synapse, or brain diseases (Supplementary Table 4). (**b**) Odd ratio of protein sets related to brain diseases (Fisher exact; Stars indicate p-values passing Bonferroni correction) (**c**) Protein-protein interaction (PPI) network for the most conserved and divergent protein coding genes associated with brain diseases. The RSPH1 protein accumulates variants with high Combined Annotation Dependent Depletion (CADD), a score that estimates the deleteriousness of a genetic variant.

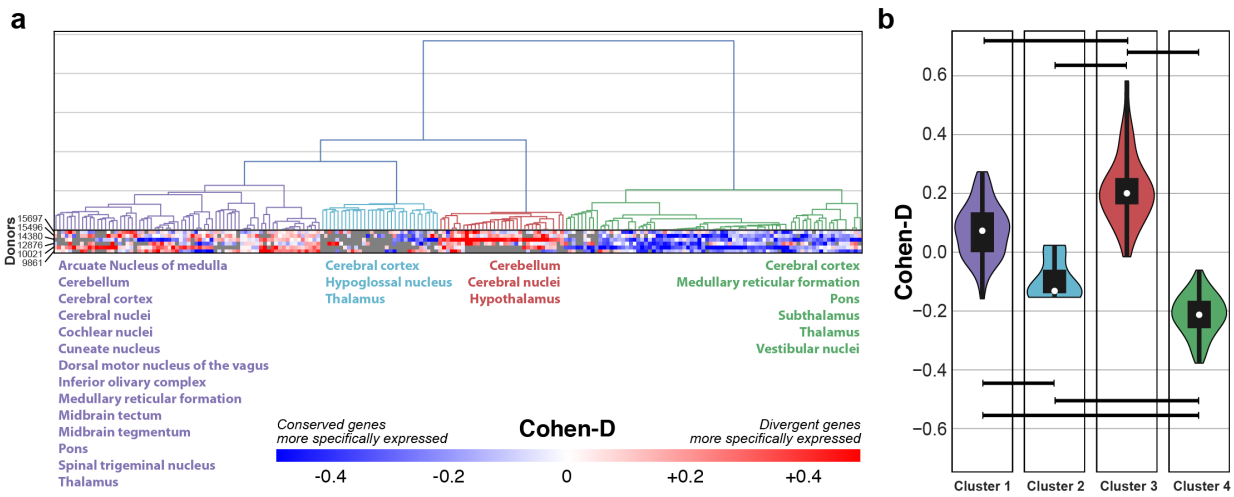

Figure 4 Differential expression signature of evolution across subgroups of cortical and subcortical brain structures. (**a**) Hierarchical clustering on the different parts of the brain from the 6 donors in the Allen Brain atlas according to the differences in expression specificity of the highly divergent and conserved proteins. Brain regions with similar evolutionary signature are indicated in the same color under each of the 4 identified clusters. The divergent proteins are more specifically expressed in the brain regions from cluster 3. (**b**) Violin plot indicating the effect sizes of the differences in expression specificity, averaged across the 6 brain donors for each brain structure in each of the 4 clusters (Kruskal-Wallis test W=134.8, p<0.001; Horizontal black bars indicate post hoc comparison with Mann-Whitney passing Bonferroni correction).

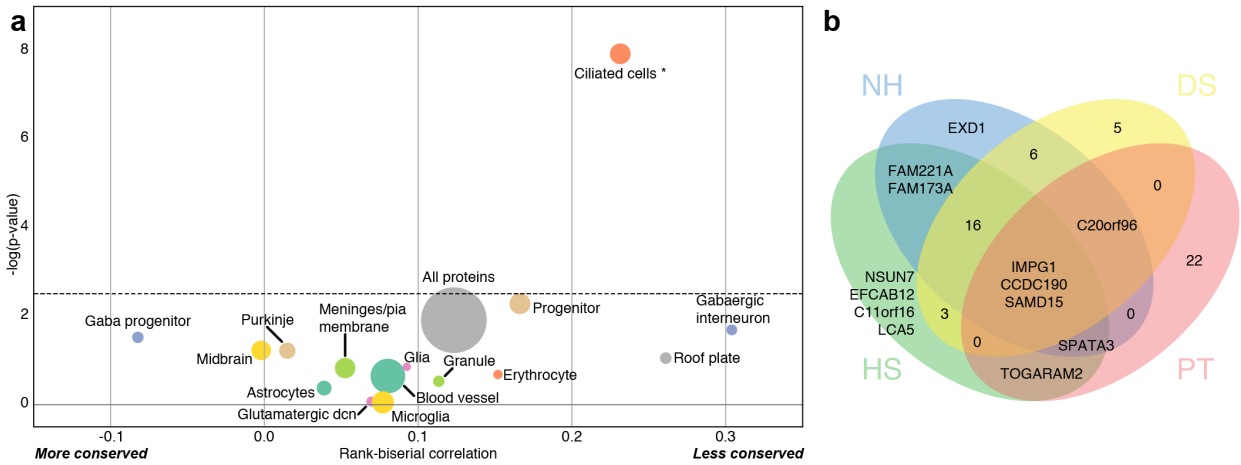

Figure 5 Evolution of protein-coding genes related to cerebellum cell types. (**a**) Funnel plot summarizing the evolution of protein-coding genes specifically expressed in different cell types of the cerebellum. (**b**) Venn diagram summarizing the divergent proteins of *Homo sapiens* (HS), Neanderthal (NH), Denisovan (DS), and *Pan troglodytes* (PT) which are specifically expressed in ciliated cells.

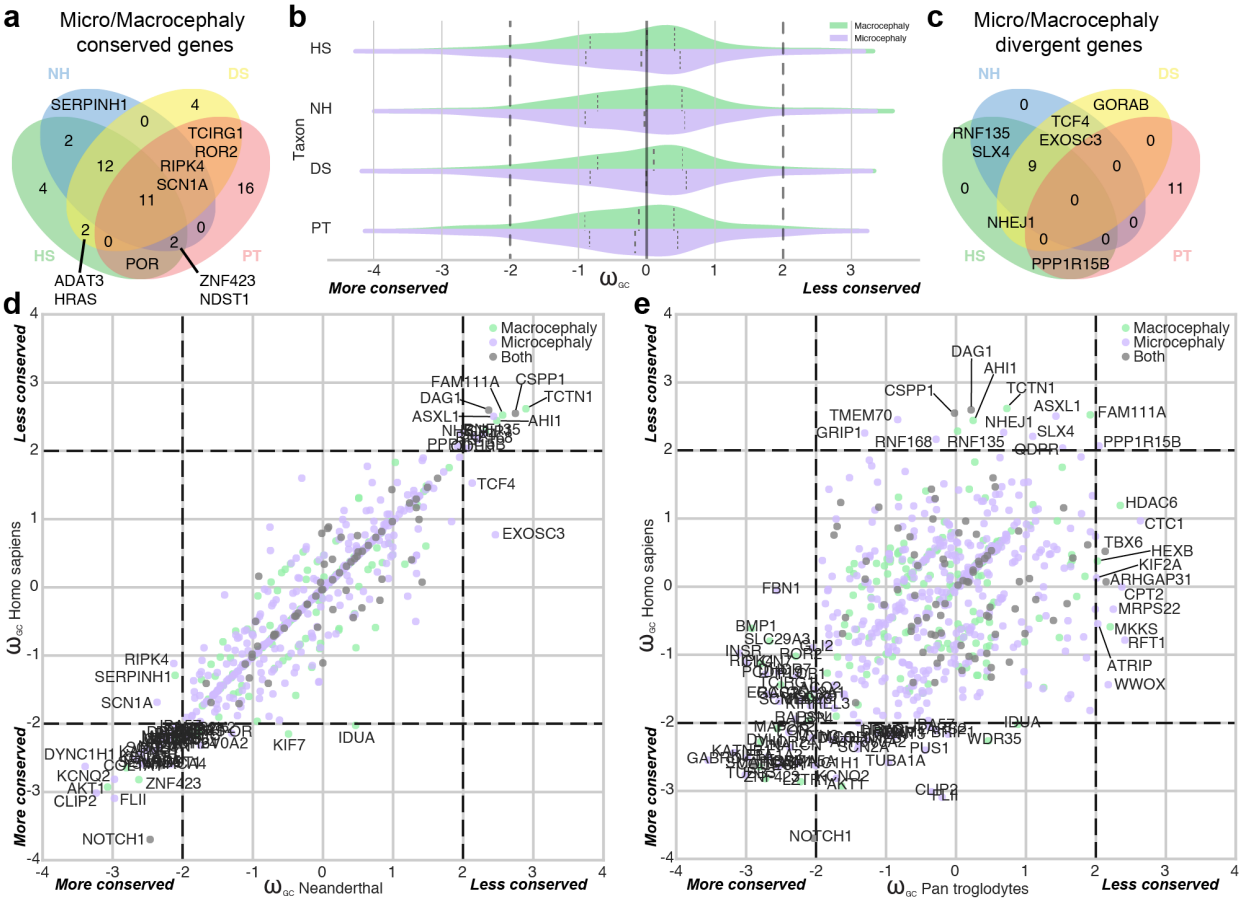

Figure 6. **Evolution of the proteins associated with micro- or macrocephaly in humans.** Comparison between ω_{GC12} across taxa for the microcephaly and macrocephaly associated genes. Venn diagrams for the conserved (**a**) and divergent (**c**) proteins for *Homo sapiens* (HS), Neanderthal (NH), Denisovan (DS), and *Pan troglodytes* (PT). (**b**) Violin plots of $\boldsymbol{\omega}_{GC}$ for proteins associated with microcephaly (purple), macrocephaly (green) or both (grey). Scatter plots of $\boldsymbol{\omega}_{GC}$ for the same proteins confronting *Homo sapiens* with either Neanderthal (**d**) or *Pan troglodytes* (**e**).

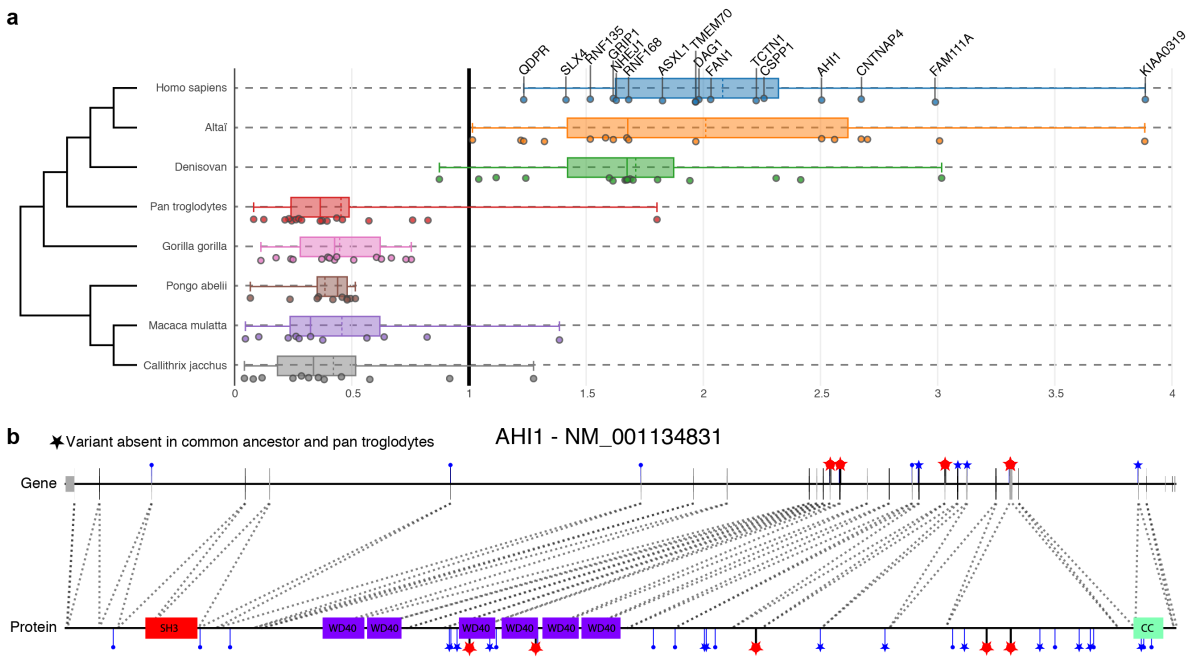

Figure 7. Examples of brain disorder-associated proteins with a specific divergence for hominins during primate evolution. (**a**) Representation of 16 proteins with dN/dS>1 in *Homo sapiens* and archaic hominins but dN/dS<1 for the rest of the primates. (**b**) Representation of hominin specific variants in the *AHI1* gene with the correspondence on the protein; notice how two variants are falling within the WP40 functional domains. Red stars indicate variants (CADD>5) present in *Homo sapiens*, Neanderthal, and Denisovan, but not in the *Pan troglodytes* compared to the Ancestor.

Methods

1. Genetic sequences

Alignments on the reference genome: We gathered and reconstructed aligned sequences on the reference human genome version hg19 (release 19, GRCh37.p13). For the primate, common ancestor sequence, we took the Ensemble 6 way Enredo-Pecan-Ortheus (EPO)[67] multiple alignments v71, related to human (hg19), chimpanzee (panTro4), gorilla (gorGor3), orangutan (ponAbe2), rhesus macaque (rheMac3), and marmoset (calJac3). For the two ancestral hominins Altai and Denisovan, we integrated variants detected by Castellano and colleagues[68] onto the standard hg19 sequence (http://cdna.eva.mpg.de/neandertal/, date of access 2014-07-03). We finally took the whole genome alignment of all the primates used in the 6-EPO from the UCSC website (http://hgdownload.soe.ucsc.edu/downloads.html, access online: August 13th, 2015).

VCF annotation: We combined the VCF file from Castellano and colleagues[68] with generated VCF files from the Ancestor and primates sequences alignment. The annotation of the global VCF was done with ANNOVAR[69] (version of June 2015) and using the data base: refGene, cytoBand, genomicSuperDups, esp6500siv2_all, 1000g2014oct_all, 1000g2014oct_afr, 1000g2014oct_eas, 1000g2014oct_eur, avsnp142, ljb26_all, gerp++elem, popfreq_max, exac03_all, exac03_afr, exac03_amr, exac03_eas, exac03_fin, exac03_nfe, exac03_oth, exac03_sas. We also used the Clinvar database (https://ncbi.nlm.nih.gov/clinvar/, date of access 2016-02-03).

2. ω_{GC12} calculus

Once all alignments were gathered, we extracted the consensus coding sequences (CCDS) of all protein-coding genes referenced in Ensembl BioMart Grc37, following HGNC (date of

access 05/05/2015) and NCBI Concensus CDS protein set (date of access 2015-08-10) and computed the number of non-synonymous mutations N, the number of synonymous mutations S, the ratio of the number of nonsynonymous mutation per non-synonymous site dN, the number of synonymous mutations per synonymous site dS, and their ratio dN/dS — also called ω—between all taxa and the ancestor using the yn00 algorithm implemented in the PamL software[70]. To avoid infinite or null results, we computed a corrected version of dN/dS, i.e. if S was null, we took it equal to one for avoiding null numerator. Finally, we obtained our ω_{GC12} measure by correcting for the GC12 content of the genes by using a general linear model and calculating a Z-score within each taxon[27]. Indeed, the GC content has been associated with biases in mutation rates, especially in primates[71] and humans[72]. We kept only the 11667 genes with one-to-one orthologs in primates (extracted for GRCh37.p13 with Ensemble Biomart, access online: February 27th, 2017).

3. Genes sets

We used different gene sets, starting at the tissue level and then focused on the brain and key pathways. For body tissues, we took the Illumina Body Map 2.0 RNA-Seq data, which consists of 16 human tissue types, including adrenal, adipose, brain, breast, colon, heart, kidney, liver, lung, lymph, ovary, prostate, skeletal muscle, testes, thyroid, and white blood cells (for more info : https://personal.broadinstitute.org/mgarber/bodymap_schroth.pdf; data pre-processed with Cufflinks, accessed the 5th of May 2015 at http://cureffi.org). We also used the microarray dataset by Su and colleagues[73] (Human U133A/GNF1H Gene Atlas, accessed the the 4th of May 2015 at http://biogps.org). Finally, we also replicated our results on the recently available RNAseq data from the GTEx Consortium[74] (https://www.gtexportal.org/home/).

For the brain, we used the Su and colleague dataset and the sets of the Human Protein Atlas (accessed the 7th of November 2017 at https://www.proteinatlas.org). For the biological pathways associated with the brain we used KEGG (accessed the 25th of February 2015, at http://www.genome.jp/kegg/), synaptic genes curated by the group of Danielle Posthuma at Vrije Universiteit (accessed the 1st of September 2014, at https://ctg.cncr.nl/software/genesets), and mass spectrometry data from Loh and colleagues[75]. Finally, for the diseases associated with the brain we combined gene sets generated from Human Phenotype Ontology (accessed the 5th of April 2016, at http://human-phenotype-ontology.github.io) and OMIM (accessed the 5th of April 2016, at https://omim.org), and curated lists: the 65 risk genes proposed by Sanders and colleagues[76] (TADA), the genes candidates for autism spectrum disorders from SFARI (accessed the 17th of July 2015 at https://gene.sfari.org), the Developmental Brain Disorder or DBD (accessed the 12th of July 2016 at https://geisingeradmi.org/care-innovation/studies/dbd-genes/), and the Cancer Census (accessed the 24th of November 2016 at cancer.sanger.ac.uk/census). Notice that the combination of both HPO & OMIM allows to be the most inclusive and not miss any potential candidate gene, however this leads to not specific association.

SynGO was gently provided by Matthijs Verhage (access date: 2019-01-11). This ontology is a consistent, evidence-based annotation of synaptic gene products developed by the SynGO consortium (2015-2017) in collaboration with the GO-consortium. It extends the existing Gene Ontology (GO) of the synapse and follow the same dichotomy between Biological Processes (BP) and Cellular Components (CC).

For single cell transcriptomics datasets, we identified the genes specifically highly expressed in each cell types, following the same strategy used for the other RNAseq datasets. The data for single-cell developing human cortex was gently provided by Maximilian Haeussler (available at https://cells.ucsc.edu; access date: 2018-10-30). The data of a single-cell transcriptional atlas of the developing murine cerebellum[40] was gently provided by Robert A. Carter (access date: 2019-01-29). For each cell type, we combined expression values observed across all available replicates to guarantee a good signal-to-noise ratio. We then calculated the associated genes in *Homo sapiens* following paralogy correspondence between Humans and Mice (Ensembl Biomart accessed on 2019-02-23).

4. Genes nomenclature

We extracted from Ensembl Biomart all the EntrezId of the protein coding genes for Grc37. We used the HGNC database to recover their symbol. For the 46 unmapped genes, we manually search on NCBI the official symbol.

5. Neanderthal introgression

The level of introgression from Neanderthal to modern humans was assessed by averaging the probabilities of Neanderthal ancestry calculated for European and Asian population for each SNP of the 1000 Genome Project dataset [51,77].

6. McDonad-Kreitman-test (MK) and Neutrality Index (NI)

To detect the fixation of variant in the *Homo sapiens* population we first computed the relative rate of non-synonymous to synonymous polymorphism (pN/pS) from the 1000 Genomes VCF for all SNPs, for SNPS with maximum allele frequency (MAF) <1% and >5%. SNPs were annotated with ANNOVAR across 1000 Genomes Project (ALL+5 ethnicity groups), ESP6500 (ALL+2 ethnicity groups), ExAC (ALL+7 ethnicity groups), and CG46 (see http://annovar.openbioinformatics.org/en/latest/user-guide/filter/#popfreqmax-and-

popfreqall-annotations for more details). We then used the McDonald–Kreitman test by calculating the Neutrality Index (NI) as the ratio of pN/pS and dN/dS values[78]. We considered the divergent genes as fixed in the population when NI < 1.

7. Protein-Protein Interaction network

To plot Protein-Protein Interaction (PPI), we combined eight human interactomes: the Human Integrated Protein-Protein Interaction rEference (HIPPIE) (accessed the 10th of August 2017 at http://cbdm-01.zdv.uni-mainz.de/~mschaefer/hippie/), the Agile Protein Interactomes DataServer (APID) (accessed the 7th os September 2017 at http://cicblade.dep.usal.es:8080/APID/), CORUM – the comprehensive resource of mammalian protein complexes (accessed the 13th of July 2017 at http://mips.helmholtz-muenchen.de/corum/), and five PPI from of the Center for Cancer Systems Biology (CCSB) (accessed the 12th of July 2016 at): four high-quality binary Protein-Protein Interactions (PPI) using a systematic primary yeast two-hybrid assay (Y2H): HI-I-05 by Rual et colleagues[79], Venkatesan-09 by Venkatesan and colleagues[80], Yu-11 by Yu and colleagues[81] and HI-II-14 by Rolland and colleagues[82], plus one high-quality binary literature dataset Lit-BM-13 by Rolland and colleagues[82], comprising all PPI that are binary and supported by at least two traceable pieces of evidence (publications and/or methods).

8. NeuroVault analyses

We used the NeuroVault website[44] to collect 19,244 brain imaging results from fMRI-BOLD studies (T and Z score maps) and their correlation with the genes expression of the Allen Brain atlas[45]. The gene expression from the Allen Brain atlas were normalized and projected into the MNI152 stereotactic space used by NeuroVault using the spatial coordinates provided by the Allen Brain institute. Because an inverse relationship between cortical and subcortical

expression dominated the pattern of expression for many genes, the correlations are calculated for cortex and subcortical structures separately.

9. Allen Brain data

We downloaded the Allen Brain atlas microarray-based gene data from the Allen Brain website (accessed the 19th of January 2018) at http://www.brain-map.org). Six adult brains had microarrays, and the right hemisphere was missing for three donors so we considered only the left one for our analyses. For each donor, we averaged probes targeting the same gene and falling in the same brain area. Then, we log-normalised the data and computed Z-scores: across the 20787 genes for each brain region to get expression levels; across the 212 brain areas for each gene to get expression specificity. For genes with more than one probe, we averaged the normalized values over all probes available. As a complementary dataset, we also used a mapping of Allen Brain Atlas on the 68 brain regions of the Freesurfer atlas[83] (accessed the 4th of April 2017 at https://figshare.com/articles/A_FreeSurfer_view_of_the_cortical_transcriptome_generated_from_the_Allen_Human_Brain_Atlas/1439749).

10. Statistics

Enrichment analyses: We first computed a two-way hierarchical clustering on the normalized dN/dS values ($\boldsymbol{\omega}_{GC}$) in different across the whole genome (see Fig1b; note: 11,667 genes were part of the analysis for a medium quality coverage for *Homo sapiens*, Neanderthal, Denisovan, and *Pan troglodytes*; see Supplementary table 2). According to 30 clustering indices[84], the best partitioning in term of evolutionary pressure was in two clusters of genes: constrained (N=4825; in HS, mea=-0.88 med=-0.80 sd=0.69) and divergent

(N=6842; in HS, mea=0.60 med=0.48 sd=0.63. For each cluster, we calculated the enrichment in biological functions in Cytoscape[85] with the BINGO plugin[86]. We took as background all the 12,400 genes. To eliminate redundancy, we first filtered all the statistically significant Gene Ontology (GO) with less than 10 genes or more than 1000, and then combine the remaining ones with the EnrichmentMap plugin[87]. We used a P-value Cutoff = 0.005, FDR Q-value Cutoff = 0.05, and Jaccard Coefficient = 0.5.

For the Cell-type Specific Expression Analysis (CSEA)[88] we used the CSEA method using the online tool http://genetics.wustl.edu/jdlab/csea-tool-2/. This method associate any gene lists with the brain expression profiles across cell type, region, and time period.

Wilcoxon and rank-biserial correlation: To measure how each gene set was significantly more conserved or divergent than expected by chance, we calculated a Wilcoxon test of the normalized dN/dS values ($\boldsymbol{\omega}_{GC}$) of the genes in the set against zero (the mean of the genome). For quantification of the effect size, we used matched pairs rank-biserial correlation as introduced by Kerby[89]. Following the calculation of non-parametric Wilcoxon signed rank tests, the rank-biserial correlation is calculated as the difference between the proportion of negative and positive ranks compared to the total sum of ranks:

$$rc = \frac{\sum r_+ - \sum r_-}{\sum r_+ + \sum r_-} = f - u$$

It corresponds to the difference between the proportion of observation favourable to the hypothesis (f) minus the proportion of observation that are unfavourable (u), thus representing an effect size. Like other correlational measures, its value ranges from minus one to plus one, with a value of zero indicating no relationship. In our case, a negative rank-biserial correlation corresponds to a gene set where more genes have negative ω_{GC} values

than positive, thus revealing more conservation than the average of all the genes (i.e. $\omega_{GC} = 0$). At contrary, a positive rank-biserial correlation corresponds to a gene set more divergent than expected. All statistics related to Figures 1d, 2a and 2b are summarized in the Supplementary table 3.

Validation by resampling technique: To correct for potential bias of the length of the coding sequence or the global specificity of the gene expression (Tau, see section A3), we also run bootstraps. For each permutation, we took randomly the same number of genes as in the sample within the total set of genes with non-missing dN/dS. We corrected for CCDS length and GC content by bootstrap resampling method. We estimated the significance of the null hypothesis rejection by calculating how many draws of bootstraps (B_i) fall below and above the observed measure (m). The related empirical p-value was calculated according to this formula:

$$p = 2 * \min\left(\frac{1 + \sum_i B_i \geq m}{N + 1}, \frac{1 + \sum_i B_i \leq m}{N + 1}\right)$$

Acknowledgements

We thank J-P. Changeux, L. Quintana-Murci, E. Patin, G. Laval, and B. Arcangioli for sharing their advices and comments, and the members of the Human Genetics and Cognitive Functions lab for helpful discussions. We also are grateful to S. Paäbo, C. Gorgolewski, R. Carter, L. Bally-Cuif, A. Chedotal, C. Berthelot, H. Roest Crollius, M. Haeussler, M. Verhage and the SynGO consortium for providing key datasets without which the current work could not have been possible. This work was supported by the Institut Pasteur; Centre National de la Recherche Scientifique; the University Paris Diderot; the Fondation pour la Recherche Médicale [DBI20141231310]; The human brain project; the Cognacq-Jay foundation; the

Bettencourt-Schueller foundation; and the Agence Nationale de la Recherche (ANR) [SynPathy]. This research was supported by the Laboratory of Excellence GENMED (Medical Genomics) grant no. ANR-10-LABX-0013, Bio-Psy and by the INCEPTION program ANR-16-CONV-0005, all managed by the ANR part of the Investment for the Future program. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Author contributions

G.D. and T.B. devised the project and conceived the main conceptual ideas. G.D. developed the methods, performed the analyses, and designed the figures. G.D. and T.B. discussed the results and wrote the manuscript. S.M. developed the companion website.

References

1. Dunbar, R. I. M. & Shultz, S. Why are there so many explanations for primate brain evolution? *Phil Trans R Soc B* **372**, 20160244 (2017).

2. Varki, A., Geschwind, D. H. & Eichler, E. E. Human uniqueness: genome interactions with environment, behaviour and culture. *Nat. Rev. Genet.* **9**, nrg2428 (2008).

3. Neubauer, S., Hublin, J.-J. & Gunz, P. The evolution of modern human brain shape. *Sci. Adv.* **4**, eaao5961 (2018).

4. Balsters, J. H. *et al.* Evolution of the cerebellar cortex: the selective expansion of prefrontal-projecting cerebellar lobules. *NeuroImage* **49**, 2045–2052 (2010).

5. Charrier, C. *et al.* Inhibition of SRGAP2 Function by Its Human-Specific Paralogs Induces Neoteny during Spine Maturation. *Cell* **149**, 923–935 (2012).

6. Florio, M. *et al.* Human-specific gene ARHGAP11B promotes basal progenitor amplification and neocortex expansion. *Sci. N. Y. NY* **347**, 1–9 (2015).

7. Suzuki, I. K. *et al.* Human-Specific NOTCH2NL Genes Expand Cortical Neurogenesis through Delta/Notch Regulation. *Cell* **173**, 1370-1384.e16 (2018).

8. Dennis, M. Y. *et al.* The evolution and population diversity of human-specific segmental duplications. *Nat. Ecol. Evol.* **1**, 69 (2017).

9. Nuttle, X. *et al.* Emergence of a Homo sapiens-specific gene family and chromosome 16p11.2 CNV susceptibility. *Nature* **536**, 205–209 (2016).

10. Dennis, M. Y. *et al.* Evolution of Human-Specific Neural SRGAP2 Genes by Incomplete Segmental Duplication. *Cell* **149**, 912–922 (2012).

11. Enard, W. *et al.* Molecular evolution of FOXP2, a gene involved in speech and language. *Nature* **418**, 869–872 (2002).

12. Montgomery, S. H., Mundy, N. I. & Barton, R. A. ASPM and mammalian brain evolution: a case study in the difficulty in making macroevolutionary inferences about gene-phenotype associations. *Proc. R. Soc. B Biol. Sci.* **281**, 20131743 (2014).

13. Enard, W. *et al.* A humanized version of Foxp2 affects cortico-basal ganglia circuits in mice. *Cell* **137**, 961–971 (2009).

14. Atkinson, E. G. *et al.* No Evidence for Recent Selection at FOXP2 among Diverse Human Populations. *Cell* **0**, (2018).

15. Miyata, T., Kuma, K., Iwabe, N. & Nikoh, N. A possible link between molecular evolution and tissue evolution demonstrated by tissue specific genes. *Idengaku Zasshi* **69**, 473–480 (1994).

16. Wang, H.-Y. *et al.* Rate of Evolution in Brain-Expressed Genes in Humans and Other Primates. *PLoS Biol* **5**, e13 (2006).

17. Tuller, T., Kupiec, M. & Ruppin, E. Evolutionary rate and gene expression across different brain regions. *Genome Biol.* **9**, R142 (2008).

18. King, M.-C. & Wilson, A. C. Evolution at two levels in humans and chimpanzees. *Science* **188**, 107–116 (1975).

19. Pollard, K. S. *et al.* An RNA gene expressed during cortical development evolved rapidly in humans. *Nature* **443**, 167–172 (2006).

20. Changeux, J.-P. Climbing Brain Levels of Organisation from Genes to Consciousness. *Trends Cogn. Sci.* **21**, 168–181 (2017).

21. Calarco, J. A. *et al.* Global analysis of alternative splicing differences between humans and chimpanzees. *Genes Dev.* **21**, 2963–2975 (2007).

22. Bakewell, M. A., Shi, P. & Zhang, J. More genes underwent positive selection in chimpanzee evolution than in human evolution. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 7489–7494 (2007).

23. Sundaram, L. *et al.* Predicting the clinical impact of human mutation with deep neural networks. *Nat. Genet.* (2018). doi:10.1038/s41588-018-0167-z

24. Havrilla, J. M., Pedersen, B. S., Layer, R. M. & Quinlan, A. R. A map of constrained coding regions in the human genome. *Nat. Genet.* **51**, 88 (2019).

25. Dorus, S. *et al.* Accelerated Evolution of Nervous System Genes in the Origin of Homo sapiens. *Cell* **119**, 1027–1040 (2004).

26. Huang, Y. *et al.* Recent Adaptive Events in Human Brain Revealed by Meta-Analysis of Positively Selected Genes. *PLoS ONE* **8**, e61280 (2013).

27. Kapheim, K. M. *et al.* Genomic signatures of evolutionary transitions from solitary to group living. *Science* **348**, 1139–1143 (2015).

28. Yang, Z. & Bielawski, J. Statistical methods for detecting molecular adaptation. *Trends Ecol. Evol.* **15**, 496–503 (2000).

29. O'Toole, Á. N., Hurst, L. D. & McLysaght, A. Faster Evolving Primate Genes Are More Likely to Duplicate. *Mol. Biol. Evol.* **35**, 107–118 (2018).

30. Schoenemann, P. T., Sheehan, M. J. & Glotzer, L. D. Prefrontal white matter volume is disproportionately larger in humans than in other primates. *Nat. Neurosci.* **8**, 242–252 (2005).

31. Frith, C. & Dolan, R. The role of the prefrontal cortex in higher cognitive functions. *Cogn. Brain Res.* **5**, 175–181 (1996).

32. Tao, X., West, A. E., Chen, W. G., Corfas, G. & Greenberg, M. E. A calcium-responsive transcription factor, CaRF, that regulates neuronal activity-dependent expression of BDNF. *Neuron* **33**, 383–395 (2002).

33. Whitney, O. *et al.* Core and region-enriched networks of behaviorally regulated genes and the singing genome. *Science* **346**, 1256780 (2014).

34. Ruano, D. *et al.* Functional Gene Group Analysis Reveals a Role of Synaptic Heterotrimeric G Proteins in Cognitive Ability. *Am. J. Hum. Genet.* **86**, 113–125 (2010).

35. Lips, E. S. *et al.* Functional gene group analysis identifies synaptic gene groups as risk factor for schizophrenia. *Mol. Psychiatry* **17**, 996–1006 (2012).

36. van Vuurden, D. G. *et al.* Pre-B-cell leukemia homeobox interacting protein 1 is overexpressed in astrocytoma and promotes tumor cell growth and migration. *Neuro-Oncol.* **16**, 946–959 (2014).

37. Hawrylycz, M. J. *et al.* An anatomically comprehensive atlas of the adult human brain transcriptome. *Nature* **489**, 391–399 (2012).

38. Franquinho, F. *et al.* The Dyslexia-susceptibility Protein KIAA0319 Inhibits Axon Growth Through Smad2 Signaling. *Cereb. Cortex N. Y. N 1991* **27**, 1732–1747 (2017).

39. Mishra, M., Inoue, N. & Heese, K. Characterizing the novel protein p33MONOX. *Mol. Cell. Biochem.* **350**, 127–134 (2011).

40. Carter, R. A. *et al.* A Single-Cell Transcriptional Atlas of the Developing Murine Cerebellum. *Curr. Biol.* **28**, 2910-2920.e2 (2018).

41. Nowakowski, T. J. *et al.* Spatiotemporal gene expression trajectories reveal developmental hierarchies of the human cortex. *Science* **358**, 1318–1323 (2017).

42. Coene, K. L. M. *et al.* OFD1 is mutated in X-linked Joubert syndrome and interacts with LCA5-encoded lebercilin. *Am. J. Hum. Genet.* **85**, 465–481 (2009).

43. Martin, C.-A. *et al.* Mutations in TOP3A Cause a Bloom Syndrome-like Disorder. *Am. J. Hum. Genet.* **103**, 221–231 (2018).

44. Gorgolewski, K. J. *et al.* NeuroVault.org: a web-based repository for collecting and sharing unthresholded statistical maps of the human brain. *Front. Neuroinformatics* **9**, 8 (2015).

45. Gorgolewski, K. J. *et al.* Tight fitting genes: finding relations between statistical maps and gene expression patterns. *F1000Research* **5**, (2014).

46. Karayannis, T. *et al.* Cntnap4 differentially contributes to GABAergic and dopaminergic synaptic transmission. *Nature* **511**, 236–240 (2014).

47. Jamain, S. *et al.* Mutations of the X-linked genes encoding neuroligins NLGN3 and NLGN4 are associated with autism. *Nat. Genet.* **34**, 27–29 (2003).

48. Abrahams, B. S. *et al.* SFARI Gene 2.0: a community-driven knowledgebase for the autism spectrum disorders (ASDs). *Mol. Autism* **4**, 36 (2013).

49. Mejias, R. *et al.* Gain-of-function glutamate receptor interacting protein 1 variants alter GluA2 recycling and surface distribution in patients with autism. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 4920–4925 (2011).

50. Paracchini, S. *et al.* The chromosome 6p22 haplotype associated with dyslexia reduces the expression of KIAA0319, a novel gene involved in neuronal migration. *Hum. Mol. Genet.* **15**, 1659–1666 (2006).

51. Sankararaman, S. *et al.* The genomic landscape of Neanderthal ancestry in present-day humans. *Nature* **507**, 354–357 (2014).

52. Dennis, M. Y. & Eichler, E. E. Human adaptation and evolution by segmental duplication. *Curr. Opin. Genet. Dev.* **41**, 44–52 (2016).

53. Hayward, P. Joubert syndrome may provide clues about human evolution. *Lancet Neurol.* **3**, 574 (2004).

54. Mekel-Bobrov, N. *et al.* Ongoing adaptive evolution of ASPM, a brain size determinant in Homo sapiens. *Science* **309**, 1720–1722 (2005).

55. Evans, P. D. *et al.* Microcephalin, a gene regulating brain size, continues to evolve adaptively in humans. *Science* **309**, 1717–1720 (2005).

56. Yu, F. *et al.* Comment on 'Ongoing adaptive evolution of ASPM, a brain size determinant in Homo sapiens'. *Science* **316**, 370 (2007).

57. Unger, S. *et al.* FAM111A mutations result in hypoparathyroidism and impaired skeletal development. *Am. J. Hum. Genet.* **92**, 990–995 (2013).

58. Hoischen, A. *et al.* De novo nonsense mutations in ASXL1 cause Bohring-Opitz syndrome. *Nat. Genet.* **43**, 729–731 (2011).

59. Ferland, R. J. *et al.* Abnormal cerebellar development and axonal decussation due to mutations in AHI1 in Joubert syndrome. *Nat. Genet.* **36**, 1008–1013 (2004).

60. Gould, D. B. & Walter, M. A. Mutational analysis of BARHL1 and BARX1 in three new patients with Joubert syndrome. *Am. J. Med. Genet. A.* **131**, 205–208 (2004).

61. Smaers, J. B., Turner, A. H., Gómez-Robles, A. & Sherwood, C. C. A cerebellar substrate for cognition evolved multiple times independently in mammals. *eLife* **7**, (2018).

62. MacLeod, C. E., Zilles, K., Schleicher, A., Rilling, J. K. & Gibson, K. R. Expansion of the neocerebellum in Hominoidea. *J. Hum. Evol.* **44**, 401–429 (2003).

63. Kochiyama, T. *et al.* Reconstructing the Neanderthal brain using computational anatomy. *Sci. Rep.* **8**, 6296 (2018).

64. Gunz, P. *et al.* Neandertal Introgression Sheds Light on Modern Human Endocranial Globularity. *Curr. Biol.* (2018). doi:10.1016/j.cub.2018.10.065

65. Koziol, L. F., Budding, D. E. & Chidekel, D. From movement to thought: executive function, embodied cognition, and the cerebellum. *Cerebellum Lond. Engl.* **11**, 505–525 (2012).

66. Rao, S. M., Mayer, A. R. & Harrington, D. L. The evolution of brain activation during temporal processing. *Nat. Neurosci.* **4**, 317–323 (2001).

67. Paten, B., Herrero, J., Beal, K., Fitzgerald, S. & Birney, E. Enredo and Pecan: Genome-wide mammalian consistency-based multiple alignment with paralogs. *Genome Res.* **18**, 1814–1828 (2008).

68. Castellano, S. *et al.* Patterns of coding variation in the complete exomes of three Neandertals. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 6666–6671 (2014).

69. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **38**, e164–e164 (2010).

70. Yang, Z. PAML 4: Phylogenetic Analysis by Maximum Likelihood. *Mol. Biol. Evol.* **24**, 1586–1591 (2007).

71. Galtier, N., Duret, L., Glémin, S. & Ranwez, V. GC-biased gene conversion promotes the fixation of deleterious amino acid changes in primates. *Trends Genet.* **25**, 1–5 (2009).

72. Kostka, D., Hubisz, M. J., Siepel, A. & Pollard, K. S. The Role of GC-Biased Gene Conversion in Shaping the Fastest Evolving Regions of the Human Genome. *Mol. Biol. Evol.* **29**, 1047–1057 (2012).

73. Su, A. I. *et al.* A gene atlas of the mouse and human protein-encoding transcriptomes. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 6062–6067 (2004).

74. Consortium, T. Gte. The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans. *Science* **348**, 648–660 (2015).

75. Loh, K. H. Proteomics: The proteomes of excitatory and inhibitory synaptic clefts. *Nat. Methods* **13**, 903–903 (2016).

76. Sanders, S. J. *et al.* Insights into Autism Spectrum Disorder Genomic Architecture and Biology from 71 Risk Loci. *Neuron* **87**, 1215–1233 (2015).

77. Deschamps, M. *et al.* Genomic Signatures of Selective Pressures and Introgression from Archaic Hominins at Human Innate Immunity Genes. *Am. J. Hum. Genet.* **98**, 5–21 (2016).

78. McDonald, J. H. & Kreitman, M. Adaptive protein evolution at the Adh locus in Drosophila. *Nature* **351**, 351652a0 (1991).

79. Rual, J.-F. *et al.* Towards a proteome-scale map of the human protein-protein interaction network. *Nature* **437**, 1173–1178 (2005).

80. Venkatesan, K. *et al.* An empirical framework for binary interactome mapping. *Nat. Methods* **6**, 83–90 (2009).

81. Yu, H. *et al.* Next-generation sequencing to generate interactome datasets. *Nat. Methods* **8**, 478–480 (2011).

82. Rolland, T. *et al.* A proteome-scale map of the human interactome network. *Cell* **159**, 1212–1226 (2014).

83. French, L. A FreeSurfer view of the cortical transcriptome generated from the Allen Human Brain Atlas. (2017). doi:10.6084/m9.figshare.1439749.v11

84. Charrad, M., Ghazzali, N., Boiteau, V. & Niknafs, A. NbClust: an R package for determining the relevant number of clusters in a data set. *J. Stat. Softw.* **61**, 1–36 (2014).

85. Shannon, P. *et al.* Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Res.* **13**, 2498–2504 (2003).

86. Maere, S., Heymans, K. & Kuiper, M. BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinforma. Oxf. Engl.* **21**, 3448–3449 (2005).

87. Merico, D., Isserlin, R., Stueker, O., Emili, A. & Bader, G. D. Enrichment Map: A Network-Based Method for Gene-Set Enrichment Visualization and Interpretation. *PLOS ONE* **5**, e13984 (2010).

88. Xu, X., Wells, A. B., O'Brien, D. R., Nehorai, A. & Dougherty, J. D. Cell Type-Specific Expression Analysis to Identify Putative Cellular Mechanisms for Neurogenetic Disorders. *J. Neurosci.* **34**, 1420–1431 (2014).

89. Kerby, D. S. The Simple Difference Formula: An Approach to Teaching Nonparametric Correlation. *Compr. Psychol.* **3**, 11.IT.3.1 (2014).