Neocortex expansion is linked to size variations in gene families with chemotaxis, cell–cell signalling and immune response functions in mammals

Atahualpa Castillo-Morales¹,²,³ Jimena Monzón-Sandoval¹,²,³,†, Alexandra A. de Sousa⁴, Araxi O. Urrutia¹,³ and Humberto Gutierrez²

¹Department of Biology and Biochemistry, University of Bath, Bath BA2 7AY, UK
²School of Life Sciences, University of Lincoln, Lincoln LN6 7TS, UK
³Milner Centre for Evolution, University of Bath, Bath BA2 7YA, UK
⁴Culture and Environment: Psychology, Bath Spa University, Bath BA2 9BN, UK

Increased brain size is thought to have played an important role in the evolution of mammals and is a highly variable trait across lineages. Variations in brain size are closely linked to corresponding variations in the size of the neocortex, a distinct mammalian evolutionary innovation. The genomic features that explain and/or accompany variations in the relative size of the neocortex remain unknown. By comparing the genomes of 28 mammalian species, we show that neocortical expansion relative to the rest of the brain is associated with variations in gene family size (GFS) of gene families that are significantly enriched in biological functions associated with chemotaxis, cell–cell signalling and immune response. Importantly, we find that previously reported GFS variations associated with increased brain size are largely accounted for by the stronger link between neocortex expansion and variations in the size of gene families. Moreover, genes within these families are more prominently expressed in the human neocortex during early compared with adult development. These results suggest that changes in GFS underlie morphological adaptations during brain evolution in mammalian lineages.

1. Introduction

Increased brain size in mammals when compared with other vertebrate taxa is thought to have played an important role in the expansion of this clade. Increased brain size during evolution has been previously related to increased behavioural complexity and the ability to cope with a changing environment [1,2]. However, the precise evolutionary drivers of brain size expansion in mammals and its relation to behavioural ability are still unclear and remain a topic of much interest and debate. This is complicated by the fact that different mammalian clades have differences in the degree of size-related changes in brain tissue [3]. Generally, large brains differ from small brains in having larger neuronal soma sizes [4], increased numbers of non-neuronal cells, in particular glia [5,6], and lower overall neuron density [7]. Large brains, however, are associated with a high metabolic cost [8–11] as well as higher demands of parental investment and delayed sexual maturation [12–16].

Brain size is a highly variable trait among mammalian and non-mammalian species with marked differences observed even between relatively close species [17–21]. Because brain size is closely associated with variations in body mass across species [22], comparative studies of brain size frequently use a corrected measure of brain size, known as encephalization index (EI), which provides a
measure of how much brain size is above (or below) what is expected for a given body size. While Ei is commonly regarded as an index that aligns more closely with behavioural capacity [1,25,24], many studies have also related behavioural complexity directly to the actual size of specific brain regions as well as to relative brain size as a whole [25–29]. Changes in relative brain size (or Ei) on the other hand, are not necessarily the result of a proportional expansion of all brain structures. In many mammalian lineages, most variations in encephalization index are closely linked to changes in the size of the neocortex [30–32], a distinctive structure of the mammalian brain and one of the most salient evolutionary innovations of the mammalian lineage [33–36].

The characteristic increase in the size of the neocortex relative to the rest of the brain has long been considered one of the primary targets of selection during mammalian brain evolution [37–39]. Increases in the absolute size of the neocortex are related to an increase in the number of functionally distinct neocortical areas [40–42], potentially allowing more complex information processing and the emergence of new behaviours [43]. In comparative studies in primates, for instance, relative size of the neocortex has been correlated with social group size [12,44,45] (but see [46]), and it has been speculated that the number of neocortical neurons may be a limiting factor in determining the number of social relationships mammals can effectively establish and manage [44]. More neocortical areas may be found in larger brains due to the lower marginal cost of devoting additional neural tissue to increasingly specialized functions [47], and an increasing number of neocortical areas may facilitate a more elaborate processing of sensory and motor information [48,49].

In the hominid lineage, the expansion of the neocortex is thought to have played a key role in the evolution of modern humans [50], including specialized areas involved in processing and production of language [51,52] as well as areas involved in identification of faces [53,54] and locations [55,56]. The neocortex in humans is widely regarded as the primary seat for the so-called higher cognitive functions, including self-awareness, consciousness, abstract reasoning and planning [57–63]. Development of the neocortex extends well into adolescence in humans and, although the structure of the layers in the neocortex is established during early prenatal development [64], the neocortex keeps growing in childhood and adolescence, reaching a peak in thickness on average at around 13 years of age, while myelination of some cortical regions can still continue after 20 years of age [65].

Despite the importance of the neocortex, the genomic features underlying its expansion during mammalian evolution remain poorly understood [30,66,67]. So far, there have been few efforts to identify features reflecting the genomic impact of brain evolution. Dorus and co-workers [68] reported accelerated sequence evolution of genes functioning in the nervous system during human origins, but this pattern was contested by later studies [69,70]. A genome-wide analysis of amino acid composition across 37 fully sequenced mammalian genomes showed that encephalization is significantly correlated with overall protein amino acid composition, although the causes of this pattern remain unclear [71].

Changes in gene family size (GFS) can reflect changes in the relative relevance of specific functions in an organism. Gene duplication events have been proposed to play a major role in the origin of novel gene functions and expression patterns [72,73]. Marked differences in GFS have been identified in Drosophila and vertebrates, with families experiencing the largest changes being enriched in distinct biological functions [74–76]. Among mammals, marked differences in the number of olfactory receptors are likely to reflect variations in the reliance of different lineages on their sense of smell [77–80]. A recent study found that encephalization in mammalian lineages is associated with significant variations in GFS, with a significant enrichment of genes associated with immune system response, chemotaxis and cell–cell signalling functions among the most positively associated gene families [81]. Here, we investigate if variations in the relative size of the neocortex or neocortex to brain size ratio (Nr) are associated with changes in GFS in mammalian lineages, and whether the extent to which any changes in GFS associated with Nr could explain previously reported associations between GFS variations and encephalization. We further explored whether any associated correlations between Nr and GFS are functionally reflected by the specific patterns of expression of Nr-associated families in the developing neocortex in humans.

2. Material and methods

2.1. Gene family annotations

Annotated gene families encompassing 28 fully sequenced mammalian genomes were obtained from Ensembl release 76 [82] (http://www.ensembl.org. Ensembl release 76). In the context of this annotation, Ensembl families are defined by clustering all Ensembl proteins along with metazoan sequences from UniProtKB. Any given gene family constitutes a group of related genes that includes both paralogues within the same species and orthologues and paralogues from other species. Any given gene can only be assigned to a single gene family. GFS in a given family for a given species was calculated as the total number of member genes contained in that gene family, for that particular species. In this study, we included all gene families with members present in at least six of the 28 mammalian species (n = 11 943). We excluded from this study any gene family with no variance in GFS across species.

2.2. Phenotype data

Body mass-corrected values of brain mass, known as encephalization index (Ei), were computed as

\[
Ei = \ln\left(\frac{\text{brain mass}}{\text{body mass}^{b}}\right).
\]

The slope (b) was estimated as 0.64 [83] based on a log–log least-squares linear regression of brain mass against body mass data from 493 mammalian species (table 1). Neocortex volumes were compiled from available literature (table 1), and include the grey and white matter of the cerebral cortex. Grey matter from paleocortical structures (entorhinal cortex, hippocortex, hippocampus and amygdala) were excluded. Nr was defined as

\[
Nr = \frac{\text{neocortex volume}}{\text{brain volume} - \text{neocortex volume}}
\]

after Dunbar [44]. Maximum lifespan (MLSP) for each species was obtained from the animal ageing and longevity database, AnAge [100]. Brain region volumes and corresponding sources as well as encephalization indexes and MLSP for all included species are shown in table 1.
2.3. Correlation coefficients of gene family size and different phenotypes

Pearson’s correlations between GFS values and the three phenotypes, Ei, Nr or MLSP, for all 11,943 gene families included in the study, were calculated using R-based statistical functions. To determine the statistical significance of any potential shift in the distribution of Pearson’s correlation coefficients when compared to random expectation, 10,000 Monte Carlo simulations of the expected distribution based on random permutations of GFS values across species were conducted and contrasted with the observed distribution of correlation coefficients using a Z-score test.

2.4. Confounding variables and phylogenetically controlled correlations

In order to remove the effect of Ei and MLSP on Nr, we calculated residuals for the multivariate regression of Nr ≈ Ei + MLSP (with Nr as the response variable and Ei and MLSP as independent covariates). For consistency, we used the exact same approach to obtain similar corrected estimates for all GFS values after correcting for any potential effects of Ei and MLSP. This was done by extracting the residuals for the multivariate regression GFS ≈ Ei + MLSP for each individual gene family. The resulting sets of residuals were then used to obtain phylogenetic independent contrasts (PIC) to further account for any effect of phylogenetic relationships on these variables [101]. The resulting independent contrasts were finally used to assess the final corrected association between Nr and GFS by simply using standard Pearson’s correlations forced through the origin. The same analysis was carried but using Nr and MLSP instead as independent covariates to generate residuals for all Ei and GFS values, from the multivariate regressions Ei ≈ Nr + MLSP and GFS ≈ Nr + MLSP respectively, followed by extraction of the corresponding PIC to assess the unbiased association between Ei and GFS (figure 1). PIC analysis was computed using the ape package in R [102]. Ultrametric phylogeny of the 28 analysed mammalian species was obtained from TimeTree database [103] (http://www.timetree.org/ . TimeTree2).

<table>
<thead>
<tr>
<th>species name</th>
<th>common name</th>
<th>non-neocortex brain volume (cm³)</th>
<th>neocortex volume (cm³)</th>
<th>ref.</th>
<th>Nr</th>
<th>Ei</th>
<th>MLSP</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Alluropoda melanoleuca</td>
<td>giant panda</td>
<td>211.80935</td>
<td>136.43571</td>
<td>[84]</td>
<td>1.81</td>
<td>−2.014</td>
<td>36.8</td>
<td></td>
</tr>
<tr>
<td>Callithrix jacobus</td>
<td>marmoset</td>
<td>7.241</td>
<td>4.371</td>
<td>[85]</td>
<td>1.52</td>
<td>−1.627</td>
<td>16.5</td>
<td></td>
</tr>
<tr>
<td>Canis familiaris</td>
<td>dog (poodle)</td>
<td>458.273</td>
<td>177.753</td>
<td>[86]</td>
<td>0.63</td>
<td>−1.699</td>
<td>24</td>
<td></td>
</tr>
<tr>
<td>Cavia porcellus</td>
<td>guinea pig</td>
<td>4.671815</td>
<td>1.5798</td>
<td>[87]</td>
<td>0.51</td>
<td>−2.948</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>Echinops telfairi</td>
<td>lesser hedgehog tenrec</td>
<td>0.566</td>
<td>0.0515</td>
<td>[85]</td>
<td>0.1</td>
<td>−3.274</td>
<td>19</td>
<td></td>
</tr>
<tr>
<td>Erinaceus europaeus</td>
<td>hedgehog</td>
<td>3.05</td>
<td>0.522</td>
<td>[85]</td>
<td>0.21</td>
<td>−2.863</td>
<td>11.7</td>
<td></td>
</tr>
<tr>
<td>Gorilla gorilla</td>
<td>gorilla</td>
<td>470.359</td>
<td>341.444</td>
<td>[85]</td>
<td>2.65</td>
<td>−1.415</td>
<td>55.4</td>
<td></td>
</tr>
<tr>
<td>Homo sapiens</td>
<td>human</td>
<td>1251.847</td>
<td>1006.525</td>
<td>[85]</td>
<td>4.1</td>
<td>0.152</td>
<td>122.5</td>
<td></td>
</tr>
<tr>
<td>Loxodonta africana</td>
<td>elephant</td>
<td>3886.7</td>
<td>2460.1</td>
<td>[88]</td>
<td>1.72</td>
<td>−1.082</td>
<td>65</td>
<td></td>
</tr>
<tr>
<td>Macaca mulatta</td>
<td>macaque</td>
<td>87.896</td>
<td>63.482</td>
<td>[85]</td>
<td>2.6</td>
<td>−1.192</td>
<td>40</td>
<td></td>
</tr>
<tr>
<td>Macropus eugenii</td>
<td>wallaby</td>
<td>11.6637</td>
<td>4.3987</td>
<td>[89]</td>
<td>0.61</td>
<td>−2.207</td>
<td>15.1</td>
<td></td>
</tr>
<tr>
<td>Microcebus murinus</td>
<td>mouse lemur</td>
<td>1.68</td>
<td>0.74</td>
<td>[85]</td>
<td>0.79</td>
<td>−1.985</td>
<td>18.2</td>
<td></td>
</tr>
<tr>
<td>Mus musculus</td>
<td>mouse (C57BL/6J)</td>
<td>0.48</td>
<td>0.12</td>
<td>[90]</td>
<td>0.32</td>
<td>−2.832</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>Mustela putorius furo</td>
<td>European polecat</td>
<td>8.8996</td>
<td>4.147</td>
<td>[91]</td>
<td>0.87</td>
<td>−2.548</td>
<td>11.1</td>
<td></td>
</tr>
<tr>
<td>Ornithorhynchus anatinius</td>
<td>platypus</td>
<td>8.57145</td>
<td>4.0992</td>
<td>[92]</td>
<td>0.92</td>
<td>−2.219</td>
<td>22.6</td>
<td></td>
</tr>
<tr>
<td>Ovis aries</td>
<td>sheep</td>
<td>100.332</td>
<td>53.793</td>
<td>[93]</td>
<td>1.16</td>
<td>−1.961</td>
<td>22.8</td>
<td></td>
</tr>
<tr>
<td>Pan troglodytes</td>
<td>chimpanzee</td>
<td>382.103</td>
<td>291.592</td>
<td>[85]</td>
<td>3.22</td>
<td>−0.948</td>
<td>59.4</td>
<td></td>
</tr>
<tr>
<td>Papio anubis</td>
<td>olive baboon</td>
<td>190.957</td>
<td>140.142</td>
<td>[85]</td>
<td>2.76</td>
<td>−1.178</td>
<td>37.5</td>
<td></td>
</tr>
<tr>
<td>Pongo abelii</td>
<td>orangutan</td>
<td>304.2</td>
<td>219.8</td>
<td>[94]</td>
<td>2.6</td>
<td>−0.892</td>
<td>59</td>
<td></td>
</tr>
<tr>
<td>Procavia capensis</td>
<td>hyrax</td>
<td>12.68</td>
<td>5.54</td>
<td>[95]</td>
<td>0.78</td>
<td>−2.255</td>
<td>14.8</td>
<td></td>
</tr>
<tr>
<td>Pteropus vampyrus</td>
<td>megabat</td>
<td>8.89</td>
<td>3.61</td>
<td>[96]</td>
<td>0.68</td>
<td>−2.204</td>
<td>20.9</td>
<td></td>
</tr>
<tr>
<td>Rattus norvegicus</td>
<td>rat</td>
<td>1.69</td>
<td>0.58</td>
<td>[95]</td>
<td>0.52</td>
<td>−2.861</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>Sarcophilus harrisii</td>
<td>Tasmanian devil</td>
<td>15.1517</td>
<td>3.7334</td>
<td>[89]</td>
<td>0.33</td>
<td>−2.792</td>
<td>13</td>
<td></td>
</tr>
<tr>
<td>Sorex araneus</td>
<td>shrew</td>
<td>0.188</td>
<td>0.0264</td>
<td>[85]</td>
<td>0.16</td>
<td>−2.832</td>
<td>3.2</td>
<td></td>
</tr>
<tr>
<td>Sus scrofa</td>
<td>pig</td>
<td>106.660</td>
<td>54.3913</td>
<td>[97]</td>
<td>1.04</td>
<td>−2.468</td>
<td>27</td>
<td></td>
</tr>
<tr>
<td>Tarsius syrichta</td>
<td>tarsier</td>
<td>3.393</td>
<td>1.768</td>
<td>[85]</td>
<td>1.09</td>
<td>−1.795</td>
<td>16</td>
<td></td>
</tr>
<tr>
<td>Tursiops truncatus</td>
<td>dolphin</td>
<td>1376.976</td>
<td>1088.615</td>
<td>[98]</td>
<td>3.78</td>
<td>−0.321</td>
<td>51.6</td>
<td></td>
</tr>
<tr>
<td>Vicugna pacos</td>
<td>alpaca</td>
<td>181.467</td>
<td>101.81</td>
<td>[99]</td>
<td>1.28</td>
<td>−1.688</td>
<td>25.8</td>
<td></td>
</tr>
</tbody>
</table>
2.5. Gene Ontology term enrichment

Gene Ontology (GO) annotations per species for biological process domains were obtained from Ensembl’s Biomart release 76 [82]. A GO term was associated with a family whenever that term was linked to any of its members in any species. To minimize the effect of very small functional categories, only terms linked to at least 200 gene families were examined (n = 116). GO terms with less than 200 gene families were assigned to a ‘small biological process GO terms’ category while gene families not annotated to any GO term in any species were grouped into a ‘not annotated’ category. Enrichment analysis of these GO terms was carried out as described in Castillo-Morales et al. [81]. Briefly, over-representation of genes associated with specific GO terms was assessed by counting the number of gene families assigned to each GO term within the analysed set of gene families. Statistical significance was numerically assessed by obtaining the expected number of families per GO in 1000 equally sized random samples derived from the overall population of gene families. Because genes vary in the number of GO terms associated with them, we adjusted for differences in the density of GO annotations between the test and background samples, by dividing the family counts per GO from each sample by the sample’s average number of GO annotations per family.

2.6. Gene expression before and after full cortical maturation

RNAseq RPKM normalized expression data summarized to genes were obtained from the NIH Transcriptional Atlas of Human Brain Development database [104] (http://brainspan.org. BrainSpan Atlas of the Developing Human Brain) for a total of 143 post-mortem human brain samples corresponding to 11 cortical regions across 13 different ages. The cortical regions include primary auditory cortex (core) (A1C), dorsolateral prefrontal cortex (DFC), postero-inferior (ventral) parietal cortex (IPC), inferolateral temporal cortex (area TEv, area 20) (ITC), primary motor cortex (area M1, area 4) (M1C), anterior (rostral) cingulate (medial prefrontal) cortex (MFC), orbitofrontal cortex (OFC), primary somatosensory cortex (area S1, areas 3,1,2) (S1C), posterior (caudal) superior temporal cortex (area TAc) (STC), primary visual cortex (striate cortex, area V1/17) (V1C) and ventrolateral prefrontal cortex (VFC). The samples covered developmental stages 16, 24, 37 post-conception weeks, four months after birth and 1, 3, 8, 13, 19, 21, 30, 36 and 37 years old. Gene expression data were further normalized against the total expression per sample, and divided into two developmental groups, corresponding to the periods before and after full maturation of cortical thickness, which occurs at about 13 years of age in humans [65]. For each gene, expression levels were averaged across stages and structures of the same developmental window and comparisons between developmental windows were carried out by means of paired Wilcoxon tests.

3. Results

In order to assess the association between gene family size, GFS, and neocortex expansion, Nr, values were compiled from the literature for 28 mammalian species with fully sequenced genomes (table 1). GFS was calculated for a total of 11,943 non-overlapping families. Pearson’s correlation coefficients between GFS and Nr were then calculated for each gene family. We found a significant over-representation of gene families with positive associations between GFS and Nr (figure 1) ($\chi^2 = 2973.263083, p < 1 \times 10^{-20}$). A Monte Carlo simulation showed that the observed shift in the distribution towards positive values is statistically significant when compared with random expectation (Z-score for observed mean $R = 2.225819868, p = 0.013$).

In order to assess whether the observed bias towards strong correlations between GFS and Nr preferentially involves gene families associated with specific biological functions (as opposed to random sets of gene families), we assessed the statistical over-representation of functional annotations (annotated GO terms per gene family, see Material and

Figure 1. Enrichment of gene family size variations (GFS) in line with increased encephalization index (Ei) and neocortex to brain size ratio (Nr) in mammals. (a) Histogram showing the distribution of correlation coefficients for GFS and Ei in 11,943 gene families encompassing 28 mammalian genomes. (b) Histogram showing the distribution of correlation coefficients for GFS and Nr in 11,943 gene families encompassing 28 mammalian genomes. In each figure, an estimation of the expected distribution derived from 10,000 Monte Carlo simulations is represented by the solid line. Inset: distribution of positive and negative correlations relative to the expected distribution (dashed line).
methods) for the 440 gene families found to be significantly associated with Ei and Nr (ð̃FDR, GFS > 0 and FDR < 0.05). A total of 18 GO functional categories were found to be significantly enriched (FDR < 0.05) among Nr-associated gene families including immune response, negative regulation of endopeptidase activity, chemotaxis, cell–cell signalling, neuropeptide signalling pathway and G-coupled receptor signalling pathway (figure 2). Notably, genes with no functional annotations showed the highest over-representation.

As Nr is known to be highly correlated to relative brain size, the observed association between Nr and changes in GFS could be explained in principle by a previously reported association between GFS and Ei or Nr, respectively (ð̃FDR, GFS > 0, FDR < 0.05, n = 440 and ð̃FDR, GFS > 0, FDR < 0.05, n = 440). The significant association between GFS and Nr residuals revealed a total of 272 families significantly associated with Nr after correction for multiple testing (phylogenetically controlled r’s > 0, FDR < 0.05; electronic supplementary material, table S1). By contrast, phylogenetically controlled correlations between equivalent GFS and Ei residuals (correcting for the effect of Nr and MLSP, see Material and methods) resulted in no gene families with a significant association after correcting for multiple testing. Nr-associated gene families after this correction against confounding variables were found to be enriched in GO terms including immune response, negative regulation of endopeptidase activity, chemotaxis, cell–cell signalling and neuropeptide signalling pathway (figure 2).

Figure 2. Gene ontology enrichment analysis of families with gene family size (GFS) variations in line with encephalization index (Ei) and neocortex to brain size ratio (Nr). Heatmap of the significance of the over-representation of GO terms (expressed as Benjamin–Hochberg (BH)-corrected p-value) among gene families most significantly associated with Ei and Nr. First two columns correspond to gene families with the most significant association between GFS and Ei or Nr, respectively (ð̃FDR, GFS > 0, FDR < 0.05, n = 440 and ð̃FDR, GFS > 0, FDR < 0.05, n = 321). Third and fourth columns represent GO terms enriched among gene families whose GFS variations display the most significant association with one of the brain phenotypes after accounting for the shared variance with the other neural phenotype, as well as the phylogenetic relationship of the analysed species using independent contrast analysis (ð̃FDR, GFS > 0, FDR < 0.05, n = 272 and ð̃FDR, GFS > 0, FDR < 0.05, n = 0 respectively). Only GO terms significantly enriched after BH multiple testing correction are shown in the figures.
If the association between GFS and Nr responds to the functional demands imposed by the development of a large neocortex, we should expect genes associated with families displaying a high correlation with Nr to also display a pronounced level of activity prior to full cortical maturation (when full cortical thickness is reached), compared with later stages. To this end, we used available gene expression data derived from human neocortex obtained from the BrainSpan Atlas of the Developing Human Brain [104] (see Material and methods). We found that gene members of this set of families showed higher expression levels during human development prior to the neocortex reaching maximum thickness (which in humans occurs around the age of 13 years) compared with later stages, reflecting a transcriptional signature of the potential involvement of some of these genes in the development of the neocortex \(p = 0.00013\).

4. Discussion

The expansion of the neocortex observed in several mammalian lineages is considered to be linked to a proliferation of new cortical areas driving increased cognitive capabilities [40,106]. The genomic drivers shaping the evolution of the brain and its morphology remain, however, poorly understood. By comparing the genomes of 28 mammalian species, here we have assessed the potential association between changes in GFS and the expansion of the neocortex. We show that neocortical expansion is indeed strongly and specifically associated with variations in GFS in mammals. Furthermore, variations in relative neocortical size account for a high proportion of the previously reported links between GFS and changes in encephalization across mammalian species [81]. This suggests that changes in GFS in line with relative brain size in mammals are actually secondary to the underlying correlation between neocortex size and encephalization. Analysis of available human neocortex gene expression data revealed that genes in families strongly and specifically associated with neocortex size variations also show significantly higher levels of expression at stages of development before (but not after) maximal cortical development is reached in humans, thereby supporting a functional role for these gene families in the ontological development of a large neocortex. Among the 272 gene families whose size was found robustly correlated with relative neocortex size, even after correcting for encephalization, MLSP and phylogenetic relationships, 16 distinct biological functions (GO terms) were found to be significantly overrepresented. Among these, cell–cell signalling and chemotaxis are known to play critical roles in the development and maintenance of the nervous system. Example of Nr-associated gene families annotated to these functions are the tyrosine kinase precursor family (ENSFM007300001521921), encoding receptor protein-tyrosine kinases and widely known to play critical roles in the regulation of cell cycle [120,121]. While a great variation in gene numbers across species has been documented in this family [122], here we report the first evidence of a strong association between these variations in this family and relative neocortex size in mammals.

Several immune-related biological functions (inflammatory response, defence response to bacteria, immune response, defence response and positive regulation of IκB kinase/NF-κB signalling) were also enriched among Nr-associated gene families. Along these lines in recent years, numerous immune-related signalling and regulatory components have also been shown to play key physiological roles in the developing and adult nervous system (for a review see [111]). This involvement of individual immune-related signalling components in neural functions has been shown to be part of a wider genetic network of immune-related molecules acting as an intrinsic component of the neural-specific regulatory machinery that ultimately shapes the normal development of the nervous system [112]. Thus, for instance, members of the tumour necrosis factor (TNF) receptor superfamily (ENSM0050000273041, a gene family found to be highly associated with neocortex expansion here), are themselves part of the extensively studied canonical pathway of activation of the transcription factor NF-κB during early development of the nervous system [111]. Interestingly, gene families with no reported functional annotations for any of its members in any species showed the highest enrichment among the gene families with the highest positive associations with relative neocortex size. Among these families, we found the neuroblastoma breakpoint gene family (ENSM00250000000879), whose members contain DUF1220 domains. DUF1220 domains have been previously linked to brain and cortical expansion in primate species, particularly in the human lineage [113,114]. Polymorphic deletions and duplications of DUF1220 domains have been associated with brain size variations in normal individuals from different human populations as well in pathological cases including microcephaly and macrocephaly [115,116]. Moreover, it has been proposed that proteins containing this domain have an important role during cortical neurogenesis, as they promote proliferation in neural stem cells [113], and during normal development they are expressed in the sub-ventricular zone precisely during the period of cortical neurogenesis [114].

Of particular importance to build a larger neocortex is the control of successive rounds of proliferation during early development, where the interplay between symmetric and asymmetric cell division is thought to be critical in shaping the particular morphology of the neocortex [117]. Consistent with this, one gene family with significant GFS changes in line with increased relative neocortex size is the ENSFM00250000003440 gene family of epithelial cell adhesion molecules, which in turn include known human developmental regulators such as EPCAM and TACSTD2. EPCAM has been shown to be involved in cell proliferation, differentiation and migration in diverse cell types [118,119] and could thus play an important role in neocortex development. A more numerous gene family found was the speedy gene family (ENSM00740001589497), which encodes proteins able to bind CDKs but having no similarity with cyclins, and some of its members are known to play a role in the regulation of cell cycle [120,121]. While a great variation in gene numbers across species has been documented in this family [122], here we report the first evidence of a strong association between these variations in this family and relative neocortex size in mammals.

5. Conclusion

In summary, we have identified a set of gene families whose sizes are positively associated with an expanded neocortex,
providing new insights into neocortex evolution. Moreover, as aberrant development and degeneration of cortical neurons has been linked with a variety of mental health pathologies and dementias [123,124], identifying genomic signatures associated with the evolution of larger brain size and neocortex expansion will critically contribute to our understanding of the molecular pathways involved in the development and maintenance of cortical areas in highly encephalized mammals including humans. As these pathways may not be present or developed to the same extent in less encephalized mammalian species, our finding could help to fill existing gaps in current knowledge gained from widely used rodent models.

## References


5. Herculano-Houzel S. 2014 The glia/neuron ratio: how it is linked with a variety of mental health pathologies and dementias [123,124], identifying genomic homologues of mammalian primary visual, auditory, and somatosensory cortex (BA1, BA2, BA3) for explaining evolutionary changes in brain size and neocortex ratio across nonhuman primates and other species. Proc. Natl Acad. Sci. USA 111, 16013–16018. (doi:10.1073/pnas.1401133111)


30. Medina L, Reiner A. 2000 Do birds possess homologues of mammalian primary visual,