HUMAN BRAIN EVOLUTION: INSIGHTS FROM MICROARRAYS

Todd M. Preuss[*‡], *Mario Cáceres*[*§], *Michael C. Oldham*[||] *and Daniel H. Geschwind*[||]

Abstract | Several recent microarray studies have compared gene-expression patterns n humans, chimpanzees and other non-human primates to identify evolutionary changes that contribute to the distinctive cognitive and behavioural characteristics of humans. These studies support the surprising conclusion that the evolution of the human brain involved an upregulation of gene expression relative to non-human primates, a finding that could be relevant to understanding human cerebral physiology and function. These results show how genetic and genomic methods can shed light on the basis of human neural and cognitive specializations, and have important implications for neuroscience, anthropology and medicine.

**Division of Neuroscience and Center for Behavioral Neuroscience, Yerkes National Primate Research Center, Emory University, 954 Gatewood Road, Atlanta, Georgia 30329, USA.*
‡Department of Pathology and Laboratory Medicine, Emory University School of Medicine, 1364 Clifton Road, Atlanta, Georgia 30322, USA.
§Department of Human Genetics, Emory University School of Medicine, 615 Michael Street, Atlanta, Georgia 30322, USA.
||Department of Neurology and Center for Neurobehavioral Genetics, The Neuropsychiatric Institute, UCLA School of Medicine, 710 Westwood Plaza, Los Angeles, California 90095, USA.
Correspondence to D.H.G. or T.M.P.
e-mails: dhg@ucla.edu; tpreuss@rmy.emory.edu
doi:10.1038/nrg1469

What distinguishes humans from other organisms — especially from our closest primate relatives, the chimpanzees (FIG. 1) — has been a question of enduring fascination to scholars and the general public alike. Some of the best-known human characteristics include bipedalism, an exceptionally large brain and unusual cognitive and behavioural abilities[1–3]. Humans also have a disease profile that differs from that of other primates[3]. For example, humans seem especially vulnerable to neurodegenerative diseases, including Alzheimer's disease (which is essentially unknown in non-human species)[4–6], HIV progression to AIDS[7] and certain epithelial cancers[8]. Finally, humans have the longest lifespan potential of any primate[9,10], a fact that is likely to be pertinent to ageing, neurodegenerative disease and cancer biology. Understanding the genetic bases of these HUMAN SPECIALIZATIONS would therefore constitute an important contribution to biomedical science[11–13], as well as to anthropology, neuroscience and psychology.

The success of the Human Genome Project and the advent of comparative genomics has greatly enhanced our ability to identify the genetic characteristics that distinguish humans from other primates and has helped us to begin to understand the genetic bases of human phenotypic specializations[1,2,11–13]. In particular, the completion of the chimpanzee genome sequence will enable us to identify the ~1–2% of nucleotide sequence differences that separate the two species[14]. The challenge now is to understand the functional consequences of the small fraction of these genetic changes that were involved in making us different from other primates. Much effort has focused on identifying genes that underwent large sequence changes or acceleration in the rate of nucleotide changes in the human lineage. This has led to the discovery that certain genes[15,16], including a substantial fraction of olfactory-receptor genes[17], have been deactivated in humans. In addition, an unusual pattern of amino-acid substitution has been found in two genes that might be related to the control of brain size[18–21], and in another gene that might be related to speech or vocal-motor control[22–24]. A recent comparison of 7,645 human and chimpanzee orthologous cDNAs[25] identified several hundred genes that have probably undergone POSITIVE SELECTION for amino-acid changes in the human lineage. More recently, comparison of the complete DNA sequences of human chromosome 21 and its homologue in the chimpanzee, chromosome 22, showed that there are amino-acid differences at 83% of the proteins encoded by the 231 genes described[14], with 20% exhibiting changes that are deemed significant enough to affect protein structure. These results indicate that changes affecting protein sequence could be far more prevalent in primate evolution than previously estimated.

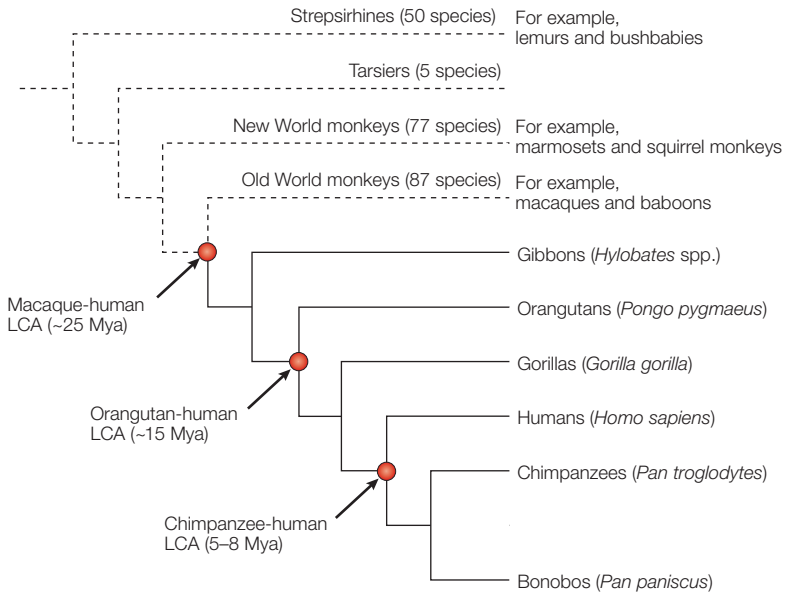

Figure 1 | **Phylogenetic relationships between humans and other primates.** Primates include over 200 extant species, which are grouped into several higher taxa. For simplicity, the diversity of most of the higher groups is collapsed into single branches; only the group to which humans belong, the primate superfamily Hominoidea (indicated by solid lines), is represented in detail here. The hominoids consist of an African great-ape group (humans, chimpanzees and gorillas), a single extant Asian great ape (the orangutan) and the gibbons. The sister group of the human lineage is the chimpanzee lineage, represented by *Pan troglodytes* (the common chimpanzee) and *Pan paniscus* (the bonobo or 'pygmy' chimpanzee). The last common ancestor (LCA) of humans and chimpanzees lived about 5–8 million years ago (Mya). Gorillas are the sister group of the human-chimpanzee group. Next, in order of relationship to humans, are the orangutans (*Pongo pygmaeus*), followed by the gibbons (genus *Hylobates*). The sister group of the hominoids is the Old World monkey group, which includes the familiar macaque monkeys, including the rhesus macaque (*Macaca mulatta*). Branching dates shown here are from REF. 76; numbers of species are from REF. 77.

The genomics revolution has made it possible to explore another dimension of evolution — changes in the temporal and spatial patterns of gene expression. The view that major phenotypic changes in evolution, especially those between humans and chimpanzees, often involve changes in gene expression has been an important theme in evolutionary biology[26–31]. The advent of DNA microarray technology allows us to quantify the expression levels of thousands of genes simultaneously and to assess thoroughly the role of gene-expression changes in evolution. By using microarrays to compare mRNA levels in samples of cerebral cortex and non-neural tissues from humans, chimpanzees and other non-human primates (TABLE 1), recent research has begun to shed light on how gene expression was modified during human evolution[32–35].

In this article, we review the results and implications of these studies, with a special focus on the gene-expression changes in the human brain. In each of these studies, the data were analysed in different ways, leading to different conclusions; taken separately, they might even be seen as contradictory. Here, we show that there is remarkable common ground in the results generated by these studies, and that clear patterns of gene-expression change in human brain evolution are beginning to emerge, providing clues about what makes our brain unique.

HUMAN SPECIALIZATION
A phenotypic characteristic of modern humans that has emerged since the divergence of the human lineage from the common ancestor of humans and chimpanzees.

POSITIVE SELECTION
A form of evolutionary change in which a mutation has a favourable effect and increases its frequency in the population at a rate greater than that predicted by neutral drift.

AGONAL STATE
The state of an individual during the time immediately preceeding death. For example, prolonged hypoxia or acidosis, can significantly affect gene expression.

OUTGROUP METHOD
Comparison of closely related species to infer the state of a common ancestor.

Cross-species microarray studies

Methodological issues. Microarray studies have been carried out using several designs, reflecting the different types of questions they have been used to address, and each of these designs poses different methodological and interpretative challenges[36,37]. In particular, nervous system studies present an extreme case of tissue heterogeneity, in which any small region of the brain might contain dozens or hundreds of cell types[38]. The ability to detect expression differences in low- or even medium-abundance RNA molecules that are restricted to specific cell populations is therefore limited[39]. However, comparative studies of different species, and especially those involving humans and non-human primates, pose additional, unique challenges. Later, we review some of the issues that these studies face, all of which highlight the importance of using alternative techniques to validate the results of microarray analysis.

Sample composition. Many of the difficulties of human microarray studies stem from the fact that the tissue samples usually have to be acquired at autopsy[36]. Ideally, subjects would be matched by age, sex, socioeconomic status, AGONAL STATE, cause of death and post-mortem delay. However, the practical obstacles to matching human samples are enormous, not least because human tissue — especially normal, control brain tissue with short post-mortem intervals — is difficult to obtain. Unlike DNA, RNA is extremely labile and, in addition to degradation, is susceptible to post-mortem artefacts[40,41], making the matter of securing suitable samples a crucial step. For example, differences in agonal state across subjects, with resultant variations in tissue pH, can contribute to the variability of microarray results and can complicate the task of detecting reliable gene-expression differences[41].

Interspecies studies that compare humans and other primates pose the further challenge of obtaining appropriate non-human primate material. Chimpanzee tissue is especially difficult to obtain because few research facilities maintain colonies of these animals. In addition, because sacrificing chimpanzees is proscribed, chimpanzee tissue (like human tissue) must be acquired post mortem. The net result is that chimpanzee samples infrequently become available for research . Studies comparing human and non-human primates face additional difficulties. For example, how do you match the ages of subjects in species with different lifespans? A 25-year-old human is a young adult, but a 25-year-old macaque is elderly. The best that can be done is to match subjects by life stage (for example, young adult). Even less tractable is the matter of matching living conditions and diet. The only solution is to remain open to the possibility that some of the interspecies differences in gene expression that are identified using microarrays might reflect such factors.

Choice of outgroups. Human specializations are, by definition, characteristics that evolved in the human lineage after it separated from the lineage leading to our closest relatives, the chimpanzees. Human–chimpanzee

OLIGONUCLEOTIDE ARRAY
A microarray made with synthetic probes, usually 25–60 bases long, each designed to hybridize to a specific mRNA. These are fabricated either *in situ* or by deposition and attachment onto a solid surface. The oligonucleotide arrays that are currently available can measure expression levels for 10,000–40,000 genes simultaneously.

INDELS
Insertions or deletions of DNA sequences in chromosomes.

NORTHERN BLOT
An experimental technique for determining the abundance and size of the transcript(s) for a particular gene in a given tissue. mRNAs are separated electrophoretically on a gel and then transferred to a membrane (blot) by capillary action. The membrane is then immersed in a labelled probe designed to hybridize to a specific mRNA.

RT-PCR
Reverse transcription PCR. Using the enzyme reverse transcriptase, RNA is converted into DNA, which is then amplified with specific primers.

comparisons by themselves, however, cannot resolve the direction of evolutionary changes among these species. A gene that is expressed in humans at a level that is five times higher than in chimpanzees might have been upregulated in humans or downregulated in chimpanzees, or might have been changed in both species. Resolving these possibilities requires knowing the level of gene expression in the last common ancestor of chimpanzees and humans, which can only be inferred indirectly using the OUTGROUP METHOD[42,43].

The best outgroups for phyletic reconstruction are the animals most closely related to the ingroups, which in the case of humans and chimpanzees are — in order of phylogenetic branching — gorillas, orangutans and gibbons (FIG. 1). Unfortunately, tissue from these species is even harder to obtain than chimpanzee tissue. The next most closely related primates are the Old World monkeys, which include macaque monkeys — animals represented by large captive populations. So, although macaques are less than ideal, practical considerations dictate that they serve as a principal source of outgroup information in microarray studies of human brain evolution.

Effect of interspecies sequence differences. The published comparative studies of gene expression in humans, apes and Old World monkeys employed Affymetrix OLIGONUCLEOTIDE MICROARRAYS that were made with human sequences (TABLE 1). However, the use of human arrays with even closely related non-human primate species is inherently problematic. Chimpanzee and macaque nucleic-acid sequences differ from humans by about 1% and 4% on average, respectively[14,44,45], excluding INDELS. Sequence mismatches reduce the affinity of an mRNA transcript for its corresponding probe sequences in the microarray, and can therefore give the false indication that the gene is expressed at lower levels in non-humans than in humans (FIG. 2a). Two factors are most likely to influence this confounding effect: the length of the array probes and the position of the mismatch within the probe[46]. This problem is especially acute with arrays that consist of short oligonucleotide probes, such as the Affymetrix arrays. A 4% divergence between humans and macaques means that, on average, each 25-nucleotide-long probe in an Affymetrix microarray will contain one nucleotide difference between the two species. Even the 1% sequence divergence of chimpanzees and humans can produce important expression-estimation errors. Moreover, the set of array probes interrogating a given gene frequently has overlapping sequences, so that a single nucleotide change can affect many, or most, of the probes at the same time. The effect of sequence divergence in human–chimpanzee comparisons is clear when gene-expression differences are checked using other techniques, such as NORTHERN BLOTTING or RT-PCR, which reveal that most false positives contain many probes with sequence mismatches between the two species[32,34].

Furthermore, sequence mismatches can interact with the NORMALIZATION procedures used in microarray analysis in a way that has not been generally appreciated.

Table 1 | **Comparative studies of human and non-human primate gene expression levels by oligonucleotide arrays**

Study (reference)	Platform	Tissue compared	Species compared*	Sex	Age (years)	Tissue origin‡	Follow-up/ validation	Data availability
Enard *et al.* (33)	HG_U95A arrays	Brain frontal cortex (area 9)	3 Humans 3 Chimpanzees 1 Orangutan	M M M	Adults Adults Adult	Autopsy Autopsy Autopsy	cDNA microarrays; northern blot; 2D protein gels	Primary data§
	HG_U95Av2 arrays	Liver	3 Humans 3 Chimpanzees 1 Orangutan	M M M	Adults Adults Adult	Surgery Autopsy Autopsy	cDNA microarrays	Primary data§
Cáceres *et al.* (32)	HG_U95Av2 arrays	Brain frontal and temporal cortex (several regions)	5 Humans 4 Chimpanzees 4 Rhesus	3 M/2 F 1 M/3 F 1 M/3 F	30–71 1–24 1–9	Surgery and autopsy Autopsy Autopsy	cDNA microarrays; RT-PCR; *in situ* hybridization	Processed data‖
	HG_U95Av2 arrays	Heart	3 Humans 2 Bonobos	2 M/1 F M	20–25 Neonates	Autopsy Autopsy		Processed data‖
Marvanova *et al.* (52)	HG_U95Av2 arrays	Brain prefrontal cortex	1 Human 1 Chimpanzee 2 Cynos 2 Marmosets	Not known M F F	Not known 7 3–3.5 2–2.5	Autopsy Autopsy Autopsy Autopsy		Processed data¶
Uddin *et al.* (35)	HG_U133A and B arrays	Brain anterior cingulate cortex (area 24)	3 Humans 1 Chimpanzee 1 Gorilla 3 Rhesus	F F M 2 M/1 F	14–28 34 41 4–11	Autopsy Autopsy Autopsy Autopsy		Primary data#
Karaman *et al.* (35)	HG_U95Av2 arrays	Fibroblast primary cell lines	13 Humans 10 Bonobos 11 Gorillas	5 M/8 F 5 M/5 F 5 M/6 F	30–79 2–20 2–35	Cell culture Cell culture Cell culture	Northern blot	Primary data**

*The number of genetically different individuals from each species analysed in each study shown (in some cases, several samples of the same individual were analysed). The following species were compared: human, *Homo sapiens*; chimpanzee, *Pan troglodytes*; bonobo (pygmy chimpanzee), *Pan paniscus*; gorilla, *Gorilla gorilla*; rhesus, *Macaca mulatta*; cynos, *Macaca fascicularis*; marmoset, *Callithrix jacchus*. ‡Post-mortem intervals for all the autopsy samples were less than 14 hours, except for the human sample in the Marvanova *et al.*[52] study, which was 19 hours. The autopsy samples of humans and great apes were obtained from individuals that died of natural causes. Primary or processed data for the arrays can be found at: §http://email.eva.mpg.de/~khaitovi/supl1/affymetrix.html; ‖http://www.teragenomics.com/public/publications_caceres01.asp; ¶http://www.uku.fi/~wong/nonhuman.htm; #http://www.genetics.wayne.edu/lgross/results/MICROARRAY_STUDIES.html; **http://hacialab.usc.edu/Supplement/karaman_etal_2003/index.html. 2D, two dimensional; M, male; F, female.

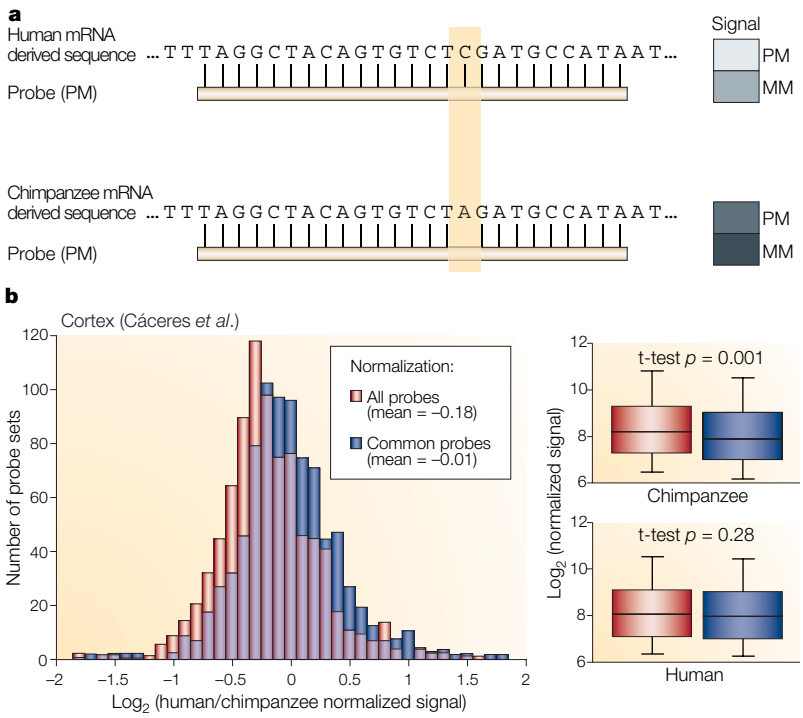

Figure 2 | **Effect of interspecific sequence differences on oligonucleotide array hybridization. a** | Hybridization of one of the oligonucleotide array probes for the catenin-α 1 gene with human and chimpanzee mRNA derived sequences. The perfect match (PM) probe is designed to be complementary to the human sequence. A mismatch (MM) probe (not shown here) is used as a control for non-specific hybridization, and differs from the PM probe by a single nucleotide change in the central position. In chimpanzees, the presence of a single C to A nucleotide change with respect to humans (yellow rectangle) markedly reduces the hybridization signal for the PM probe to similar levels of those of the MM probe (signal strength is shown in the panels on the right; a lighter shade in this panel correlates with a stronger signal). **b** | Effect of normalization on the array signal for 838 genes for which probes have identical sequences between humans and chimpanzees. Left diagram, histogram of the normalized signal in humans divided by that in chimpanzees for each gene (base-2 logarithm transformed). When normalization is carried out with all the probes on the array (represented in red), the expression levels of chimpanzee genes that do not include sequence changes with respect to humans tend to be overestimated, resulting in a predominance of negative values for the base-2 logarithm of the human/chimpanzee ratio (mean = –0.18). Using only the array probes that are common to both species (represented in blue) eliminates this effect and yields a distribution of the human/chimpanzee ratio centered around 0 (mean = –0.01). Right diagram, box plots show significantly higher normalized signals in chimpanzees after normalization using all array probes ($p = 0.001$), but this is not seen for humans ($p = 0.28$). Results shown correspond to the average of each species for the cerebral cortex data from REF. 32.

When comparing array results from many samples, each of which is run on a different array, the signal intensities in each array are normalized to compensate for differences in initial RNA amounts and labelling and detection efficiencies[47]. When non-human primate samples are hybridized on human arrays, we would expect them to show lower average hybridization efficiencies and signal intensities. In the presence of sequence mismatches, however, normalization has the effect of over-emphasizing the average signal levels for the non-human primates. This results in the overestimation of expression levels of all the genes that do not show sequence changes with respect to humans, and can give the appearance of general upregulation in non-human primates (FIG. 2b). Fortunately, as will be discussed later, there are ways to identify probes that have interspecific mismatches, so that analyses can take these factors into account. Alternatively, arrays made with longer oligonucleotide probes, or cDNA MICROARRAYS, should be less sensitive to small sequence differences across species, and therefore, in principle, should be more suitable for comparative studies.

NORMALIZATION
Mathematical processing of raw data to reduce the effects of variables introduced by the experimental design or method used. For microarrays, such variables might include differences in fluorescent dye incorporation, the amount of cRNA or cDNA hybridized to the array, hybridization conditions or the arrays themselves.

Confirmation of microarray results. The methodological difficulties that have been outlined indicate that reliable conclusions about evolutionary changes in gene expression require microarray results to be validated by additional investigation. Confirmatory evidence could include replications of the microarray studies with samples from additional individuals; use of different array techniques that are less sensitive to sequence differences, such as cDNA microarrays; or examination of expression changes in single genes by northern blotting, QUANTITATIVE REAL-TIME RT-PCR, or *in situ* hybridization[48,49]. Furthermore, as mRNA and protein levels are not always correlated[50,51], determining the biological significance of mRNA-level differences ultimately requires parallel studies of changes in protein expression, which can be examined by western blotting, immunohistochemistry and high-throughput proteomic techniques[49] that include analysis of post-translational modifications.

Human and chimpanzee gene-expression changes

We focus on five microarray studies of gene-expression differences in humans and non-human primates: four that deal primarily with the brain[32,33,35,52], and in some cases also with other tissues, and one that deals with fibroblasts[34]. All of the studies used similar human oligonucleotide arrays for the initial quantification of mRNA expression levels; the main features of their experimental designs and sample characteristics are summarized in TABLE 1. Three additional papers[53–55] have re-examined the results of the first empirical study, published by Enard and colleagues[33]. Surprisingly, the two most comprehensive brain studies, which include the largest samples of humans and chimpanzees, and validation of results by independent methods[32,33], arrived at seemingly different conclusions. As we go on to discuss, however, the results from these studies are, in fact, similar and can be used to make several conclusions and interpretations about gene-expression changes during the evolution of the human brain. First, the rate of gene-expression changes in the brain accelerated during human evolution. Second, gene-expression changes in the evolution of the human brain primarily involved increased expression (upregulation). Third, approximately 2–4% of genes expressed in the cerebral cortex seem to have different levels of expression in humans and chimpanzees, whereas a larger number of differences are detected in other non-neural tissues, such as those of the heart and liver.

In this section, we review these conclusions and their possible implications in detail. Furthermore, we will discuss a recent paper by Uddin and colleagues[35] that challenges the asymmetry of gene-expression patterns in the human brain that is apparent in the other main data sets. Finally, we consider possible criticisms derived

from the general methodological issues mentioned previously — especially the use of human arrays for interspecies comparisons — and provide new supporting evidence based on additional information that is now available, such as the chimpanzee genome sequence. An additional study on regional differences in gene expression in human and chimpanzee brains[56] became available while this paper was in review, and its results are largely consistent with those described here.

Accelerated gene-expression change in human brain evolution. In the first published account of gene-expression differences between humans and chimpanzees, Enard and colleagues[33] concluded that the human brain underwent more pronounced changes in gene expression than those that occurred in the chimpanzee lineage. In their paper[33], the evolutionary changes in expression were characterized using a DISTANCE METRIC. The authors concluded that there was an acceleration of gene-expression change in human brain evolution compared with that in chimpanzees. Since then, their data have been analysed in other ways, and patterns consistent with their initial conclusions have been noted (FIG. 3a). For example, Gu and Gu[53] tallied the number of genes that showed significant expression changes in each lineage and found statistical evidence that, in the brain, this number was approximately three times higher in humans than in chimpanzees. Using more conventional types of distance metrics than those used by Enard and colleagues, both Hsieh *et al.*[54] and Gu and Gu[55] built phylogenetic trees on the basis of expression differences, and obtained a much longer branch corresponding to the human brain than to the chimpanzee brain. Additional support for an increased number of expression changes in the human brain comes from the other two available data sets[32,35], although in both cases, the asymmetry between the human and chimpanzee lineages seems to be less than in the Enard *et al.*[33] data. By contrast, in all the previous analyses, gene-expression changes in the liver accumulated at comparable rates in the human and chimpanzee lineages (FIG. 3a).

The characterization of human brain evolution by Enard *et al.*[33] in terms of an acceleration of gene-expression change has led some to conclude that there are more genes that are differentially expressed between humans and chimpanzees in the brain than in other tissues (for example, see REF. 57). In fact, the data sets from both Enard and colleagues[33] (as re-analysed by Hsieh *et al.*[54] and Gu and Gu[55]) and Cáceres and colleagues[32] indicate that approximately twice as many genes exhibit expression differences in the liver than in the brain, and Cáceres and colleagues[32] also reported more differences in the heart than in the brain.

Upregulation of gene expression in the human brain. Another important observation common to these studies is the presence of an excess of genes with higher expression levels in the brain of humans compared with that of chimpanzees. In both the data sets from Enard *et al.*[33] and Cáceres *et al.*[32], 66–95% of the genes that are differentially expressed in human and

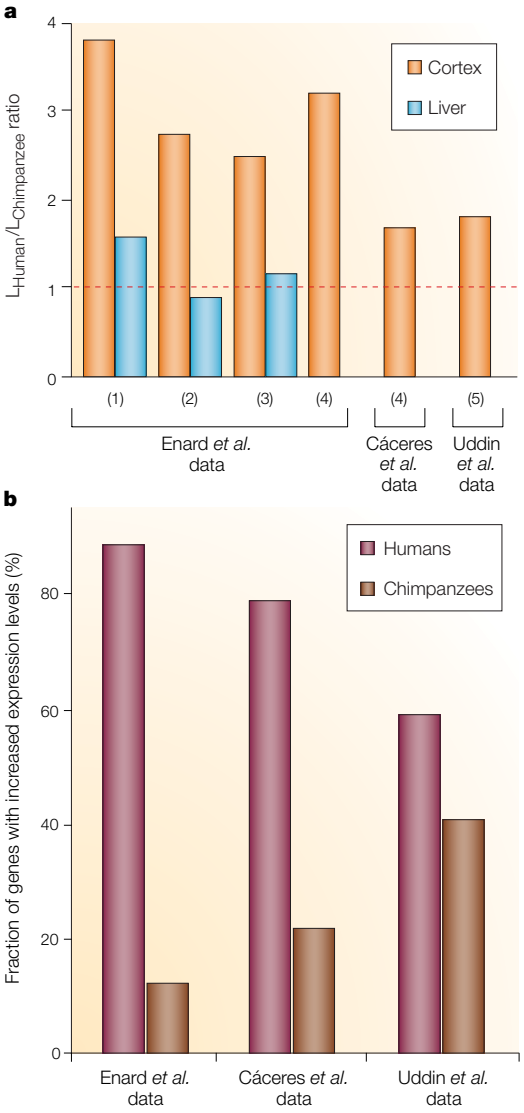

Figure 3 | **Patterns of gene-expression changes between human and chimpanzee brains. a** | Acceleration of gene-expression changes in the human lineage. Ratio of expression changes specific to the human lineage (L_{Human}) and chimpanzee lineage ($L_{Chimpanzee}$) in cerebral cortex (orange) and liver (blue) for analyses from three published data sets. The dashed red line indicates a ratio equal to 1, corresponding to a similar degree of expression change in each lineage. For cerebral cortex, the change in the human lineage is always more pronounced than in the chimpanzee lineage, whereas this is not the case for the liver. Numbers in parentheses correspond to analyses carried out in the following references: (1) Enard *et al.*[33]; (2) Gu and Gu[53]; (3) Hsieh *et al.*[54]; (4) Cáceres *et al.*[32]; (5) Uddin *et al.*[35], with results taken from three published data sets (Enard *et al.*, Cáceres *et al.* and Uddin *et al.*). When different criteria were used in the analyses, only the results from the most stringent analysis are shown. **b** | Upregulation of gene expression in the human brain. The percentage of genes differentially expressed in the cerebral cortex that show increased levels in humans (purple) is significantly higher than that in chimpanzees (brown) for the three published data sets, indicating that there are higher expression levels for many genes in the human brain. The human–chimpanzee comparison was performed according to the BullFrog analysis described in Cáceres *et al.*[32]. (In all three published data sets, sign-test *p*-value <0.001.)

cDNA MICROARRAY
A microarray made by deposition of gene-specific, PCR-amplified inserts from cDNA clones, which can be from several hundred to several thousand bases long. cDNA arrays typically measure expression levels for 5,000–30,000 genes.

QUANTITATIVE REAL-TIME RT-PCR
A procedure in which DNA amplification in a PCR reaction is measured during its log-linear phase by monitoring the accumulating signal that is provided by a fluorescent dye or gene-specific fluorescent probe incorporated into the PCR product.

DISTANCE METRIC
A measure of similarity or dissimilarity that can be used to organize groups according to their degree of relation to one another. For example, the Euclidian distance metric that distinguishes two genes, or groups of genes, is often defined as the square root of the sum of their squared expression differences.

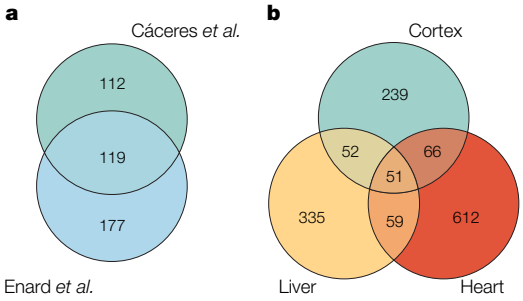

Figure 4 | **Agreement between different human–chimpanzee comparisons using oligonucleotide microarrays.** **a** | Comparison of the number of genes differentially expressed in human and chimpanzee cerebral cortices according to the Cáceres *et al.*[32] and Enard *et al.*[33] data. Approximately half the genes identified in each independent data set are common to both studies. **b** | Venn diagram of genes differentially expressed in several human and chimpanzee tissues. The number shown for the cerebral cortex corresponds to the genes that are found to be differentially expressed in either the Cáceres *et al.*[32] or Enard *et al.*[33] data sets. Fifty-one genes show expression differences between humans and chimpanzees in all tissues, and each pair of tissues share a similar number of genes that are differentially expressed. Results shown were obtained from REF. 32.

chimpanzee cerebral cortices show increased expression levels in humans, independently of how the data were analysed[32,54] (FIG. 3b). This bias is also obvious when considering gene-expression changes in the human or chimpanzee lineage. Using the orangutan[33] or rhesus macaque[32] as an outgroup, between 69% and 92% of the genes with expression changes in the human brain were upregulated[32,53]. Conversely, gene-expression analysis of liver and heart samples showed no evidence of expression asymmetry, with a similar proportion of genes increased to those decreased in both humans and chimpanzees[32,53,54]. A trend towards a greater proportion of genes that are upregulated in humans was detected in the examination of multiple fibroblast lines from humans and non-human primates[34], although it disappeared when the probes containing sequence differences between humans and chimpanzees were eliminated. Therefore, the pervasive gene-expression upregulation in humans is not a general feature of human–chimpanzee microarray comparisons, but rather a brain-specific phenomenon.

Although both the data sets from Enard *et al.*[33] and Cáceres *et al.*[32] provide evidence for predominant upregulation in human brain evolution, Uddin and colleagues[35] reported a 2:1 ratio of downregulated to upregulated genes in the human lineage. However, the results of this study must be interpreted with caution for several reasons. First, samples were small, with only one chimpanzee and one gorilla included. Second, the observed human downregulation could be explained, in part, by the focus on many expression changes of small magnitude. As discussed above, the compensatory effect of array normalization tends to overestimate expression levels of genes in non-human primates that have no or few sequence differences with respect to humans (FIG. 2b). A similar rise in the fraction of downregulated genes was observed by Gu and Gu[53] and Hsieh *et al.*[54] when significance levels were relaxed. When the data from Uddin *et al.*[35] were re-analysed using the same criteria employed in other studies (for example, those by Cáceres *et al.*[32]), 59% of the genes that were differentially expressed showed higher expression levels in humans than in chimpanzees (FIG. 3b). So, when analysed in a consistent manner, using standard metrics for determining differential expression, three independent data sets provide support for the predominant upregulation of gene expression in the human brain. This is a notable result, given the differences between the data sets in the cortical regions examined, the species studied and the microarray sets employed (TABLE 1). Moreover, the upregulation hypothesis has been validated using other techniques in addition to microarrays.

How many genes are differentially expressed? As simple as this question might seem, there is no consensus on this issue, nor will there be for some time. The number of genes identified as being differentially expressed in the cerebral cortex of humans and chimpanzees depends on the analytical methods used and the number of genes interrogated by the arrays, which ranges from ~10,000 (U95A or U95Av2 arrays) to ~33,000 (U133A and U133B arrays). Published estimates currently vary between 91, based on the most stringent statistical criteria employed by Hsieh *et al.*[54], to 4,159 in the analysis by Uddin *et al.*[35], although the latter is probably overestimated owing to the inclusion of genes that differ in expression by only a small amount in a small number of samples. The results of Enard and colleagues[33] (as re-analysed by Gu and Gu[53] and Hsieh and colleagues[54]) and of Cáceres and co-workers[32] indicate that about 100 genes (or ~2% of the genes detected by the arrays) underwent expression changes during human brain evolution. The real number is certainly higher, given that the total number of genes in the human genome is estimated to be around 20,000–30,000, and that the small sample size used in each study limits the power to detect genes that show small, but real, differences in expression. Furthermore, the cellular heterogeneity of neural tissue makes it almost certain that expression differences in low-abundance genes were not detected in the analyses[38]. Therefore, current estimates should be viewed as a lower limit, and more regional and cell-type specific analyses are needed to accurately determine the magnitude of expression changes that are related to human brain evolution.

It is noteworthy, given the differences in sample composition and brain regions examined, that there is remarkably good agreement between the two most comprehensive data sets[32,33] with respect to the identity of genes that are differentially expressed in humans and chimpanzees. Cáceres and co-workers[32] found that about half of the genes identified as being differentially expressed in the cerebral cortex were common to both data sets, and most of the remaining genes showed qualitatively similar expression differences (FIG. 4a).

The correlation of the human–chimpanzee fold-change between the two data sets for all genes that are differentially expressed in humans or chimpanzees was 0.78 ($p<0.0001$). Moreover, there is also some concordance between the genes that are differentially expressed in the brain and in other tissues: 51 genes showed expression differences between humans and chimpanzees in all the tissues compared, and each pair of tissues shared a similar number of differentially expressed genes (FIG. 4b). This highlights the large degree of correlation between microarray studies performed in different laboratories using different tissue specimens.

Are comparative microarray studies reliable and valid? Any claim of acceleration of gene-expression changes or predominant upregulation in the human lineage faces the obvious objection that this finding is precisely the artefact one would expect when non-human species are interrogated using probes that are based on human sequences. Although it has been suggested that interspecies sequence differences do not affect array results[35], empirical experience indicates that they can, and do so to a considerable extent. Nevertheless, the fact that non-neural tissues exhibit approximately equal numbers of upregulated and downregulated genes indicates that the predominance of upregulated genes in human brain evolution is not entirely an artefact of sequence differences among species. Furthermore, two groups, Cáceres *et al.*[32] and Hsieh *et al.*[54], used algorithms that flag those probes with hybridization levels that are inconsistent with the levels shown by other probes for the same gene, presumably owing to sequence differences. In both cases, the re-analysis of the array data after removing these probes detected some false positives that had no real expression differences between humans and chimpanzees; however, removing these probes did not markedly affect the overall conclusions.

These corrective measures are not perfect, however, and a better approach would be to identify all of the problematic array probes by directly comparing human and chimpanzee sequences. This is now possible, as a result of the Chimpanzee Genome Project. Approximately 25% of the array probes exhibit sequence differences, and when these were excluded, re-analysis of the data sets from Enard and colleagues[33], Cáceres and colleagues[32] and Uddin and co-workers[35] supports the conclusion that there is a greater number of upregulated genes in the human cerebral cortex compared with that of chimpanzees, whereas the proportion of genes with increased levels in humans was close to 50% in all the other tissues (liver, heart or fibroblasts; M.C. and M.C.O., unpublished data).

Enard *et al.*[33] and Cáceres *et al.*[32] also used several independent methods to validate the oligonucleotide array results, including custom cDNA microarrays[32,33], northern blotting[33], quantitative RT-PCR[32] and *in situ* hybridization[32]. The results obtained with the cDNA arrays were qualitatively similar to those obtained with the oligonucleotide arrays[32,33]. Furthermore, in these studies, more than 80% of the gene-expression changes evaluated by northern blotting or quantitative RT-PCR were confirmed and, overall, more than 60 expression differences in the brains of humans and chimpanzees were validated by at least one independent method[32,33]. This degree of confirmation is similar to that obtained for human and bonobo fibroblasts in the comprehensive study of non-neural tissue by Karaman and co-workers[34].

GENE ONTOLOGY
A framework for classifying gene products hierarchically in three dimensions according to the biological process in which they are involved, the molecular function that they perform and the cellular component in which they are located.

Functional implications of expression changes

Published gene-expression studies in primates have focused more on understanding the pattern of evolutionary change (acceleration versus upregulation) than the functional implications of the changes in expression. As a first step in addressing this issue, Cáceres and colleagues[32] used the GENE-ONTOLOGY classification[58] to identify functional groups of genes that are upregulated or downregulated in the human brain. An excess of genes included in the categories of cell growth and/or maintenance and metabolism was found, especially for those involved in lipid and RNA metabolism, and the same was true for genes coding for molecular chaperones, including heat-shock proteins. A gene-by-gene survey indicates that many of the genes that are upregulated in humans code for proteins that are involved in neuronal functions and synaptic activity[32,54]. This is consistent with the gene-ontology analysis of Uddin and co-workers[35], which identified expression changes of numerous genes related to aerobic energy metabolism and neuronal function in the human lineage, together with transcription, translation and mRNA-processing genes. TABLE 2 summarizes the gene-expression changes in the human brain that are common to array data taken from Enard *et al.*[33] and Cáceres *et al.*[32]. TABLE 2 also illustrates some of the functional categories described above, as well as information on the human diseases that mutations in these genes cause.

The elevated expression levels of many genes in the human brain, and their relationship to energy metabolism, indicates that the general level of physiological activity in the adult cerebral cortex might be higher in humans than in chimpanzees[32,35]. Additionally, the upregulation in gene expression could be the result of an accumulation of mRNA molecules in human cells to allow for rapid responses to external stimuli. The human brain might also have adapted to the damaging effects of maintaining high rates of neural activity over the course of a long lifetime by increasing the expression of chaperones and other genes with neuroprotective functions[32]. It is tempting to speculate that it is the gradual decline of these protective mechanisms during a long lifetime that causes our increased susceptibility to neurodegenerative diseases. However, the expression changes in genes involved in mRNA processing and translation[32,35] indicate that additional studies are needed to ascertain whether the higher amounts of mRNA yield corresponding increases in protein levels in the human brain.

A programme for future research

Refinement and validation of expression changes. How can we begin to put the long list of putative gene-expression changes in the human brain into a

Table 2 | **Genes with the most consistent expression changes in the human brain in microarray studies**

Gene	Fold change*	Validation‡	Protein function‖	Human disorder (OMIM reference number)¶
Upregulated genes				
SMAD1	5.1–6.4	RT-PCR	Regulation of transcription/signal transduction	
GTF2I	2.5–4.2	RT-PCR	Regulation of transcription/signal transduction	Williams–Beuren syndrome (194050)
CROC4	2.8–2.9	cDNA arrays	Regulation of transcription?	
C21orf33	5.3		Regulation of transcription?	
ZFP36L2	4.1		Regulation of transcription/cell proliferation	
PMS2L5	3.0		DNA repair/mismatch repair	
SF3A3	10.5		RNA processing/RNA splicing	
RGL1	3.1		Signal transduction	
PDE4DIP	9.3		Signal transduction?	
ENTPD6	2.1	cDNA arrays	Nucleotide metabolism/hydrolase activity	
CA2	11.5	RT-PCR, *in situ* hybridization	One-carbon compound metabolism/carbonate dehydratase activity	Osteopetrosis with renal tubular acidosis/ marble brain disease (259730)
NAGPA	5.2	cDNA arrays	Carbohydrate metabolism/protein modification	
GM2A	10.2		Lipid metabolism/sphingolipid catabolism	Tay–Sachs disease, AB variant (272750)
SPTLC1	18.0		Lipid metabolism/sphingolipid metabolism	Hereditary sensory neuropathy type I (162400)
PRDX6	2.9		Lipid metabolism/response to oxidative stress	
OSBPL8	9.6		Lipid transport/steroid metabolism	
GOSR1	12.4	RT-PCR§	Intracellular protein transport	
HSPA2	5.2–9.4	RT-PCR	Chaperone activity/heat-shock protein activity	
COL6A1	12.5	cDNA arrays RT-PCR	Cell adhesion/extracellular matrix	Bethlem myopathy (158810)
THBS4	5.3	RT-PCR	Cell adhesion and motility/extracellular matrix	Premature coronary heart disease (600715)
WIRE	9.6		Actin cytoskeleton organization?	
Downregulated genes				
TWIST1	–6.7	RT-PCR, *in situ* hybridization	Regulation of transcription/skeletal development	Saethre–Chotzen syndrome (101400)
DDX17	–10.8	RT-PCR	RNA processing/RNA helicase activity	
ACADSB	–3.0		Lipid metabolism/energy pathways	2-Methylbutyrylglycinuria (600301)

*Average fold change in gene expression levels between human and chimpanzee cortex in the Enard *et al.*[33] and Cáceres *et al.*[32] oligonucleotide array data. Positive and negative values correspond to genes with higher and lower expression in humans, respectively. When two different probe sets of the same gene show significant expression differences, both values are shown. ‡Independent confirmation by Cáceres *et al.*[32] of expression differences in human cortex using real-time RT-PCR, cDNA microarray analysis and *in situ* hybridization. §M.C., unpublished results. ‖Function of the protein encoded by the gene according to the information available in the SOURCE (source.stanford.edu), LocusLink (www.ncbi.nlm.nih.gov/LocusLink), OMIM (www.ncbi.nlm.nih.gov/OMIM) and AmiGO (www.godatabase.org) databases. Inferred functions are indicated by question marks. ¶Human disorders that are associated with mutations affecting the gene listed can be found in the OMIM (www.ncbi.nlm.nih.gov/OMIM) database. The OMIM reference numbers are in brackets. RT-PCR, A type of PCR in which RNA is converted into DNA, which is then amplified.

functional context and prioritize genes for further study? First, it is essential to extend the confirmatory work that has already been done[32] to cover more genes, using independent samples, where possible, at the mRNA and protein levels. Second, there is a need for more accurate information about expression levels in the outgroup species, without which the expression changes that are specific to the human lineage cannot be reliably identified. Rhesus macaques are a logical choice for additional investigation, given the availability of tissue and the fact that the genome of this species is currently being sequenced. Differences between human and macaque sequences will, however, continue to complicate the quantification of macaque gene expression using human microarrays, making it necessary to validate results and use array designs based on longer oligonucleotides or cDNAs. Third, studies with additional samples and brain regions, and combined analysis of published data sets, will help to overcome the difficulties that are inherent in analysing small numbers of samples.

Phenotype discovery. One of the main attractions of comparative genetic studies of humans and non-human primates is the use of genetic differences to guide the characterization of previously unrecognized human phenotypic specializations. Apart from the obvious fact that the volume of the human brain is, on average, about three times greater than that of the chimpanzee brain, only a few differences in internal brain structure have been documented (reviewed in REF. 59). Moreover, there is little consensus regarding which characteristics should be considered as species-specific cognitive specializations (for example, see REFS 60–66). Using techniques such as *in situ* hybridization and immunohistochemistry to localize the mRNAs and proteins that are

encoded by differentially expressed genes, species differences in microarray results could guide us to the discovery of differences in cell composition, structure and function of brain tissue. For example, the human anterior cingulate cortex contains much greater numbers of a particular type of pyramidal cell — the so-called spindle cells — than in other primates, and human spindle cells are also unusually large[67]. The disproportionate representation of spindle cells in the human cingulate cortex could result in a disproportionate representation of the genes expressed in them, and might therefore contribute to the differences in microarray results reported in humans compared with other primates[35].

It is also important to consider that although the human cerebral cortex is far larger than that of the chimpanzee, it contains only perhaps 50% more neurons. As comparative gene-expression studies are performed using similar amounts of RNA from each species, and RNA abundance is relatively constant per gram of brain tissue, a several-fold increase in the expression of a neuron-specific gene might actually reflect a much larger fold-change per human neuron than per chimpanzee neuron. This could have several consequences, including failure to detect expression changes in specific neuronal populations in humans. This known difference in the density of a principal cell type highlights both the potentially profound effects of confounding factors related to studying a highly heterogeneous tissue type such as the brain, and the need to put changes in gene expression in their appropriate anatomical context using other techniques.

Additionally, many of the most profound changes in human brain evolution have probably involved changes in the expression of genes that function early in development, when the fundamental structural divisions of the brain and cerebral cortical regions are specified. Documenting these changes requires carrying out comparative studies of fetal brain material from humans, chimpanzees and other primates, a project that poses practical obstacles. It is possible, however, that differences in adult gene-expression levels will provide clues to the underlying developmental processes that lead to the adult phenotype. For example, several of the genes that show expression changes in the human brain are involved in development and neurogenesis[32] or are associated with neurodevelopmental syndromes in humans[68] (TABLE 2).

The biggest challenge will be to translate these differences in gene expression into functional insights into key human specializations — such as language, higher cognitive abilities and a long lifespan — and a better understanding of why humans are so vulnerable to neurodegenerative disease[1–3]. This will, no doubt, require molecular-anatomical investigations, but will more immediately benefit from metanalyses and modelling approaches that assess higher-order relationships among genes (that is, gene-ontology[69,70] and NETWORK ANALYSES[71]). Furthermore, human cells in culture and *in vitro* assays could prove useful for empirically determining the functional consequences of gene-expression changes.

Mechanisms and causation. Evolution can lead to changes in gene expression in many ways, including modifications in the coding sequences of *trans*-acting transcription factors, DNA methylation changes, nucleotide substitutions, small insertions or deletions in promoters and other *cis*-acting regulatory regions, transposable element insertions, gene duplications and chromosomal rearrangements. Furthermore, microarray results could reflect species differences in splicing and in the exon composition of mRNAs. Large amounts of microarray expression data give us the opportunity to consider regulatory networks and to identify key regulators that could be driving evolutionary changes in gene expression.

The completion of the chimpanzee genome allows examination of the effect of sequence changes on gene-expression levels. In the recent comparison of high-quality sequence from chimpanzee chromosome 22 and its homologue, human chromosome 21, no obvious differences were found in the upstream regions of the genes that were identified as being differentially expressed[14], but more detailed analysis with a larger number of genes is needed. There is, however, a statistical association between genes with higher levels of human–chimpanzee expression divergence and regions of human duplication and chromosomal rearrangement[72], but the full extent to which these differences underlie gene-expression changes remains to be determined. Another preliminary study has found a significant number of differences in DNA-methylation patterns between humans and chimpanzees in the brain[73]. Understanding the unique contributions of these different mechanisms to changes in gene expression will be fundamental to understanding the evolution of human neural and cognitive specializations; it seems likely that these expression changes will reflect multiple types of regulatory change.

Another key question is whether these expression changes reflect evolutionary adaptations that have been shaped by natural selection, or whether they are neutral changes that reflect the random accumulation of mutations. Khaitovich and colleagues[74] suggested that most expression differences are neutral or nearly neutral. Similarly, Hsieh and co-workers[54] could only attribute a small percentage of the expression differences between humans and chimpanzees to DIVERSIFYING SELECTION. Unfortunately, there is no test for the effects of positive selection on gene expression that is equivalent to those used to assess nucleotide-sequence variation. Moreover, it is not clear how neutral evolution could account for the bias towards upregulation in human brain evolution, when other tissues show no such bias. A deeper understanding of the regulatory changes responsible for expression differences, together with the analysis of intraspecific patterns of nucleotide variation in these regions, should make it possible to better assess the role of natural selection in the modulation of gene expression, and should also help us to identify the changes that are most likely to be of functional significance for the evolution of human neural and cognitive specializations.

NETWORK ANALYSIS
Analysis of the individual interactions between constituents, which, when grouped together, describe a network. In the case of gene-expression data, network analysis entails the identification of relationships among genes or groups of genes across different experimental conditions or tissue samples.

DIVERSIFYING SELECTION
Natural selection against the mean value of a quantitative trait, therefore favouring individuals at the two tails of the phenotypic distribution.

Concluding remarks

Microarray analyses of gene-expression differences in humans and chimpanzees have allowed researchers to begin uncovering some of the changes that characterize human brain evolution at the molecular level, including the upregulation of many genes. Connecting these data to the critical phenotypes of interest, such as the emergence of language in humans, theory of mind and our particular susceptibility to certain neurological diseases, will require careful gene-by-gene research into the structural and functional context of the neural systems that underlie our remarkable human qualities. For example, understanding human speech development in the context of other primates might benefit from comparative studies with other species that are vocal learners, such as songbirds[24,75]. Similar comparative analyses that combine DNA-sequence and gene-expression data could prove useful in unravelling the genetic basis of other uniquely human features, such as the skeletal and morphological changes involved in bipedal locomotion. Gene-expression studies that are focused on multiple, early developmental stages, and many tissues, such as bone, muscle, brain and spinal cord, will probably be necessary. Additional primate resources, including tissue banks and cDNA libraries from several species, will increase our ability to more efficiently approach the question of how we became human — a question that remains daunting, but no longer beyond our reach.

1. Carroll, S. B. Genetics and the making of *Homo sapiens*. *Nature* **422**, 849–857 (2003).
2. Hacia, J. G. Genome of the apes. *Trends Genet.* **17**, 637–645 (2001).
3. Gagneux, P. & Varki, A. Genetic differences between humans and great apes. *Mol. Phylogenet. Evol.* **18**, 2–13 (2001).
4. Gearing, M., Tigges, J., Mori, H. & Mirra, S. S. β-Amyloid (Aβ) deposition in the brains of aged orangutans. *Neurobiol. Aging* **18**, 139–146 (1997).
5. Gearing, M., Rebeck, G. W., Hyman, B. T., Tigges, J. & Mirra, S. S. Neuropathology and apolipoprotein E profile of aged chimpanzees: implications for Alzheimer disease. *Proc. Natl Acad. Sci. USA* **91**, 9382–9386 (1994).
6. Walker, L. C. & Cork, L. C. in *Alzheimer's Disease* 2nd edn (eds Terry, R. D., Katzman, R., Bick, K. L. & Sisodia, S. S.) 233–243 (Lippincott Williams & Wilkins, Philadelphia, 1999).
7. Rutjens, E. *et al.* Lentivirus infections and mechanisms of disease resistance in chimpanzees. *Front Biosci.* **8**, d1134–1145 (2003).
8. McClure, H. M. Tumors in nonhuman primates: observations during a six-year period in the Yerkes primate center colony. *Am. J. Phys. Anthropol.* **38**, 425–429 (1973).
9. Blurton Jones, N. G., Hawkes, K. & O'Connell, J. F. Antiquity of postreproductive life: are there modern impacts on hunter-gatherer postreproductive life spans? *Am. J. Human Biol.* **14**, 184–205 (2002).
10. Hawkes, K., O'Connell, J. F., Jones, N. G., Alvarez, H. & Charnov, E. L. Grandmothering, menopause, and the evolution of human life histories. *Proc. Natl Acad. Sci. USA* **95**, 1336–1339 (1998).
11. McConkey, E. H. & Varki, A. A primate genome project deserves high priority. *Science* **289**, 1295–1296 (2000).
12. Olson, M. V. & Varki, A. Sequencing the chimpanzee genome: insights into human evolution and disease. *Nature Rev. Genet.* **4**, 20–28 (2003).
13. Varki, A. A chimpanzee genome project is a biomedical imperative. *Genome Res.* **10**, 1065–1070 (2000).
14. Watanabe, H. *et al.* DNA sequence and comparative analysis of chimpanzee chromosome 22. *Nature* **429**, 382–388 (2004).
The first comparison of the whole sequence of a human chromosome and its chimpanzee homologue, identifying numerous insertions and deletions, and changes in most of the encoded proteins.
15. Chou, H. H. *et al.* A mutation in human CMP-sialic acid hydroxylase occurred after the *Homo-Pan* divergence. *Proc. Natl Acad. Sci. USA* **95**, 11751–11756 (1998).
16. Stedman, H. H. *et al.* Myosin gene mutation correlates with anatomical changes in the human lineage. *Nature* **428**, 415–418 (2004).
17. Gilad, Y., Man, O., Paabo, S. & Lancet, D. Human specific loss of olfactory receptor genes. *Proc. Natl Acad. Sci. USA* **100**, 3324–3327 (2003).
18. Zhang, J. Evolution of the human ASPM gene, a major determinant of brain size. *Genetics* **165**, 2063–2070 (2003).
19. Evans, P. D. *et al.* Adaptive evolution of ASPM, a major determinant of cerebral cortical size in humans. *Hum. Mol. Genet.* **13**, 489–494 (2004).
20. Kouprina, N. *et al.* Accelerated evolution of the ASPM gene controlling brain size begins prior to human brain expansion. *PLoS Biol* **2**, E126 (2004).
21. Evans, P. D., Anderson, J. R., Vallender, E. J., Choi, S. S. & Lahn, B. T. Reconstructing the evolutionary history of microcephalin, a gene controlling human brain size. *Hum. Mol. Genet.* **13**, 1139–1145 (2004).
22. Enard, W. *et al.* Molecular evolution of *FOXP2*, a gene involved in speech and language. *Nature* **418**, 869–872 (2002).
A study of the molecular evolution of the first gene related to a human speech and language disorder; the encoded protein contains two human-specific amino acids in positions that are conserved in all other species.
23. Lai, C. S., Fisher, S. E., Hurst, J. A., Vargha-Khadem, F. & Monaco, A. P. A forkhead-domain gene is mutated in a severe speech and language disorder. *Nature* **413**, 519–523 (2001).
24. Teramitsu, I., Kudo, L. C., London, S. E., Geschwind, D. H. & White, S. A. Parallel *FoxP1* and *FoxP2* expression in songbird and human brain predicts functional interaction. *J. Neurosci.* **24**, 3152–3163 (2004).
25. Clark, A. G. *et al.* Inferring nonneutral evolution from human-chimp-mouse orthologous gene trios. *Science* **302**, 1960–1963 (2003).
The first large-scale comparison of human, chimpanzee and mouse coding sequences, identifying hundreds of genes that probably underwent selective changes during human evolution.
26. King, M. C. & Wilson, A. C. Evolution at two levels in humans and chimpanzees. *Science* **188**, 107–116 (1975).
A classic paper in human evolutionary biology, in which the authors argue that phenotypic differences between humans and chimpanzees mainly reflect changes in the regulation of gene expression.
27. De Beer, G. *Embryos and Ancestors* (The Clarendon Press, Oxford, 1940).
28. Gould, S. J. *Ontogeny and Phylogeny* (Belknap, Harvard Univ. Press, US, 1977).
29. Bonner, J. T. & Dawid, I. (eds). *Evolution and Development: Report of the Dahlem Workshop on Evolution and Development* (Springer, Berlin; New York, 1982).
30. Raff, R. A. *The Shape of Life: Genes, Development, and the Evolution of Animal Form* (Univ. Chicago Press, US, 1996).
31. Carroll, S. B., Grenier, J. K. & Weatherbee, S. D. *From DNA to Diversity: Molecular Genetics and the Evolution of Animal Design* (Blackwell Science, US, 2001).
32. Cáceres, M. *et al.* Elevated gene expression levels distinguish human from non-human primate brains. *Proc. Natl Acad. Sci. USA* **100**, 13030–13035 (2003).
The second independent microarray comparison of the human and chimpanzee cerebral cortices, which provides strong support for upregulation of gene-expression changes during human brain evolution.
33. Enard, W. *et al.* Intra- and interspecific variation in primate gene expression patterns. *Science* **296**, 340–343 (2002).
The first microarray study of gene-expression patterns in primates and the first to propose that gene-expression changes accelerated during human brain evolution.
34. Karaman, M. W. *et al.* Comparative analysis of gene-expression patterns in human and African great ape cultured fibroblasts. *Genome Res.* **13**, 1619–1630 (2003).
This microarray study compares fibroblasts among a large number of humans and apes, and provides strong evidence for the confounding effect of interspecies sequence differences in microarray results.
35. Uddin, M. *et al.* Sister grouping of chimpanzees and humans as revealed by genome-wide phylogenetic analysis of brain-gene expression profiles. *Proc. Natl Acad. Sci. USA* **101**, 2957–2962 (2004).
36. Mirnics, K. & Pevsner, J. Progress in the use of microarray technology to study the neurobiology of disease. *Nature Neurosci.* **7**, 434–439 (2004).
37. Karsten, S. & Geschwind, D. in *Current Protocols in Neuroscience* (eds Crawley, J. *et al.*) *Suppl. 20, Section 4: Unit 4. 28.* (John Wiley & Sons, Inc., New York, 2002).
38. Geschwind, D. H. Mice, microarrays, and the genetic diversity of the brain. *Proc. Natl Acad. Sci. USA* **97**, 10676–10678 (2000).
An early discussion of the key issues, such as tissue heterogeneity, that complicate studies of gene expression in the mammalian brain.
39. Luo, L. *et al.* Gene expression profiles of laser-captured adjacent neuronal subtypes. *Nature Med.* **5**, 117–122 (1999).
40. Bahn, S. *et al.* Gene expression profiling in the post-mortem human brain — no cause for dismay. *J. Chem. Neuroanat.* **22**, 79–94 (2001).
41. Li, J. Z. *et al.* Systematic changes in gene expression in postmortem human brains associated with tissue pH and terminal medical conditions. *Hum. Mol. Genet.* **13**, 609–616 (2004).
42. Hillis, D. M., Moritz, C. & Mable, B. K. *Molecular Systematics* (Sinauer Associates, Sunderland, Massachusetts, 1996).
43. Felsenstein, J. *Inferring Phylogenies* (Sinauer Associates, Sunderland, Massachusetts, 2004).
44. Fujiyama, A. *et al.* Construction and analysis of a human-chimpanzee comparative clone map. *Science* **295**, 131–134 (2002).
45. Thomas, J. W. *et al.* Comparative analyses of multi-species sequences from targeted genomic regions. *Nature* **424**, 788–793 (2003).
46. Chee, M. *et al.* Accessing genetic information with high-density DNA arrays. *Science* **274**, 610–614 (1996).
47. Quackenbush, J. Computational analysis of microarray data. *Nature Rev. Genet.* **2**, 418–427 (2001).
A useful review of methods for the normalization and analysis of microarray data.
48. Geschwind, D. H. *et al.* A genetic analysis of neural progenitor differentiation. *Neuron* **29**, 325–339 (2001).
49. Chuaqui, R. F. *et al.* Post-analysis follow-up and validation of microarray experiments. *Nature Genet.* **32** (Suppl.), 509–514 (2002).
50. Griffin, T. J. *et al.* Complementary profiling of gene expression at the transcriptome and proteome levels in *Saccharomyces cerevisiae*. *Mol. Cell Proteomics* **1**, 323–333 (2002).
51. Washburn, M. P. *et al.* Protein pathway and complex clustering of correlated mRNA and protein expression analyses in *Saccharomyces cerevisiae*. *Proc. Natl Acad. Sci. USA* **100**, 3107–3112 (2003).
52. Marvanova, M. *et al.* Microarray analysis of nonhuman primates: validation of experimental models in neurological disorders. *FASEB J.* **17**, 929–931 (2003).
53. Gu, J. & Gu, X. Induced gene expression in human brain after the split from chimpanzee. *Trends Genet.* **19**, 63–65 (2003).
A statistical re-analysis of the results from reference 33 and the first to argue for the predominance of upregulation of gene expression during human brain evolution.
54. Hsieh, W. P., Chu, T. M., Wolfinger, R. D. & Gibson, G. Mixed-model reanalysis of primate data suggests tissue and species biases in oligonucleotide-based gene expression profiles. *Genetics* **165**, 747–757 (2003).
55. Gu, J. & Gu, X. Further statistical analysis for genome-wide expression evolution in primate brain/liver/fibroblast tissues. *Hum. Genomics* **1**, 247–254 (2004).

56. Khaitovich, P. *et al.* Regional patterns of gene expression in human and chimpanzee brains. *Genome Res.* **14**, 1462–1473 (2004).
The first extensive regional analysis of gene expression in human and chimpanzee brains at one developmental stage (adult).
57. Pennisi, E. Primate evolution. Gene activity clocks brain's fast evolution. *Science* **296**, 233–235 (2002).
58. Ashburner, M. *et al.* Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nature Genet.* **25**, 25–29 (2000).
59. Preuss, T. M. in *The Cognitive Neurosciences* 3rd edn (ed. Gazzaniga, M. S.) (MIT Press, Cambridge, Massachusetts, in the press).
Reviews current evidence regarding evolutionary specializations of the human brain.
60. Cavalieri, P. & Singer, P. *The Great Ape Project: Equality Beyond Humanity* (St. Martin's, New York, 1994).
61. de Waal, F. B. M. *Chimpanzee Politics: Power and Sex Among Apes* (Harper & Row, New York, 1982).
62. Povinelli, D. J. *Folk Physics for Apes: the Chimpanzee's Theory of How the World Works* (Oxford Univ. Press, Oxford; New York, 2000).
63. Povinelli, D. J. & Preuss, T. M. Theory of mind: evolutionary history of a cognitive specialization. *Trends Neurosci.* **18**, 418–424 (1995).
64. Savage-Rumbaugh, E. S. & Lewin, R. *Kanzi: the Ape at the Brink of the Human Mind* (Wiley, New York, 1994).
65. Tomasello, M., Call, J. & Hare, B. Chimpanzees understand psychological states — the question is which ones and to what extent. *Trends Cogn. Sci.* **7**, 153–156 (2003).
66. Wallman, J. *Aping Language* (Cambridge Univ. Press, Cambridge; New York, 1992).
67. Nimchinsky, E. A. *et al.* A neuronal morphologic type unique to humans and great apes. *Proc. Natl Acad. Sci. USA* **96**, 5268–5273 (1999).
68. Morris, C. A. *et al. GTF2I* hemizygosity implicated in mental retardation in Williams syndrome: genotype-phenotype analysis of five families with deletions in the Williams syndrome region. *Am. J. Med. Genet.* **123A**, 45–59 (2003).
69. Doniger, S. W. *et al.* MAPPFinder: using Gene Ontology and GenMAPP to create a global gene-expression profile from microarray data. *Genome Biol.* **4**, R7 (2003).
70. Hosack, D. A., Dennis, G. Jr, Sherman, B. T., Lane, H. C. & Lempicki, R. A. Identifying biological themes within lists of genes with EASE. *Genome Biol.* **4**, R70 (2003).
71. Segal, E. *et al.* Module networks: identifying regulatory modules and their condition-specific regulators from gene expression data. *Nature Genet.* **34**, 166–176 (2003).
72. Marques-Bonet, T. *et al.* Chromosomal rearrangements and the genomic distribution of gene-expression divergence in humans and chimpanzees. *Trends Genet.* **20**, 524–528 (2004).
73. Enard, W. *et al.* Differences in DNA methylation patterns between humans and chimpanzees. *Curr. Biol.* **14**, R148–149 (2004).
74. Khaitovich, P. *et al.* A neutral model of transcriptome evolution. *PLoS Biol* **2**, E132 (2004).
75. Haesler, S. *et al. FoxP2* expression in avian vocal learners and non-learners. *J. Neurosci.* **24**, 3164–3175 (2004).
76. Stewart, C. B. & Disotell, T. R. Primate evolution — in and out of Africa. *Curr. Biol.* **8**, R582–588 (1998).
77. Fleagle, J. G. *Primate Adaptation and Evolution* (Academic Press, San Diego, 1998).

Acknowledgements

T.M.P. and M.C. contributed equally to this work. We thank the different authors of the primate microarray studies for making their data sets publicly available, James Thomas for the comparison of the array probes to the chimpanzee genome sequence, and David Kornack for his insightful comments. We would also like to thank the James S. McDonnell Foundation for their support of our research through a Collaborative Activities Grant (T.P., D.H.G.) and acknowledge support from the National Institute of Mental Health (D.H.G.).

Competing interests statement

The authors declare no competing financial interests.

Online links

DATABASES

The following terms in this article are linked online to:

OMIM: http://www.ncbi.nlm.nih.gov/Omim/
Alzheimer's disease

FURTHER INFORMATION

Affymetrix GeneChip Files:
http://email.eva.mpg.de/~khaitovi/supplement1.html
Affymetrix GeneChip Microarray Data Files:
http://www.genetics.wayne.edu/lgross/results/MICROARRAY_STUDIES.html
Affymetrix GeneChips in TeraGenomics:
http://www.teragenomics.com/public/publications_caceres01.asp
Chimpanzee Genome Project :
http://genome.wustl.edu/projects/chimp/
Functional Genomics and Bioinformatics:
http://www.uku.fi/~wong/nonhuman.htm
Human Genome Project:
http://www.ornl.gov/sci/techresources/Human_Genome/home.shtml
Joseph Hacia's Laboratory — Karaman Supplemental Data:
http://hacialab.usc.edu/Supplement/karaman_etal_2003/index.html
Access to this interactive links box is free online.