Evolution of primate gene expression

Philipp Khaitovich, Wolfgang Enard, Michael Lachmann and Svante Pääbo

Abstract | It has been suggested that evolutionary changes in gene expression account for most phenotypic differences between species, in particular between humans and apes. What general rules can be described governing expression evolution? We find that a neutral model where negative selection and divergence time are the major factors is a useful null hypothesis for both transcriptome and genome evolution. Two tissues that stand out with regard to gene expression are the testes, where positive selection has exerted a substantial influence in both humans and chimpanzees, and the brain, where gene expression has changed less than in other organs but acceleration might have occurred in human ancestors.

Negative selection
Removal of genetic variants in a population that decrease the fitness of their carrier. If negative selection acts on a phenotypic trait, this is also called stabilizing selection.

Positive selection
Increase in frequency of a genetic variant or a phenotypic trait because it increases the fitness of its carrier. If positive selection acts on a phenotypic trait, this is also called directional selection.

Max Planck Institute for Evolutionary Anthropology, Deutscher Platz 6, D-04103 Leipzig, Germany. Correspondence to S.P. e-mail: paabo@eva.mpg.de
doi:10.1038/nrg1940

During evolution, variation originates as mutations at the level of the genotype, whereas it is at the level of phenotype that negative selection weeds out deleterious mutations and that positive selection increases the frequency of advantageous mutations. Mutations can alter the phenotype by changing primary coding sequences of proteins and other gene products, or by changing regulatory DNA sequences that control transcription, translation or transcript degradation. The large effect of gene expression differences on the phenotype is evident from the range of cell types seen in a single organism, which all share the same genome. For this reason, it has been proposed that changes in gene expression account for most phenotypic differences between species, and in particular between humans and our closest living relatives, chimpanzees, gorillas and orangutans[1]. More than 30 years after this suggestion was first made, it is now possible to begin to test its validity by studying relative expression levels in humans and other primates on a genome-wide scale. Here we review such studies of transcriptomes and consider how combining them with genome analyses might provide general rules that govern expression evolution, and identify the molecular mechanisms that underlie human phenotypic features.

Human genome evolution

Among the 'major transitions in evolution' described by John Maynard Smith and Eörs Szathmáry (among which are the emergence of DNA, cells, eukaryotes, multicellular organisms and social insects), the most recent is the emergence of language and human culture[2]. Although the other great apes, as well as more distant primate relatives such as macaques, can use and even produce rudimentary tools[3–8], and although some of them transmit skills between individuals that vary between populations — phenomena that have been described as 'cultures'[9,10] or 'traditions'[11] — only humans seem to have developed tools or customs so complex that they cannot be reinvented within one generation[12]. The ultimate challenge for human evolutionary genetics is to understand the genetic underpinnings of this unique capacity that has emerged since we shared a common ancestor with chimpanzees 5 to 7 million years ago[13].

The first question one might ask is how special the human genome is compared with the genomes of other primates. The human and chimpanzee genomes differ in 1.23% of the bases that can be aligned between species[14]. In addition to these approximately 35 million point mutations, other differences include approximately 5 million insertions, deletions, duplications and inversions[14]. These rearrangements vary in scale from only a few nucleotides to whole chromosomes[15]. However, from our knowledge of other primates and other mammals, these are exactly the differences that one would expect for a pair of primate species that share a common ancestor 5 to 7 million years ago. So, at the level of genomic sequence, there is no easily identifiable major evolutionary transition. This might seem obvious to some and discouraging to others. However, we emphasize it because it shows that human evolution did not require any genetic changes that are qualitatively different from those expected given the current knowledge of molecular evolution.

These observations are best explained by the neutral theory of molecular evolution postulated by Motoo Kimura in the late 1960s (REF. 16) (BOX 1). According to this theory, the rate of evolutionary change in genomes is largely determined by the mutation rate and the extent to which mutations are weeded out by negative selection. The changes that are driven to fixation because of their survival or reproduction advantages are relatively rare compared with the large number of neutral changes.

Box 1 | **The neutral theory of evolution**

In the 1950s–1960s doubts began to emerge about the functional importance of every single DNA sequence mutation. It became apparent that DNA contains much more variation than could be maintained through balancing selection[87]. It was also realized that too many DNA sequence changes have accumulated since the cambrian explosion for every base to have been fixed by positive selection while deleterious mutations were being weeded out by negative selection[88,89]. In 1968, Motoo Kimura proposed the neutral theory of molecular evolution. This landmark theory postulated that the vast majority of DNA sequence substitutions observed both within and between species have no effect on the phenotype of an organism and are evolutionarily neutral[16,90]. The initial assumption of the theory is that the majority of nucleotides can be divided into two types: those that are under strong negative selection, and almost never change, and those that are neutral, and change through random evolutionary drift. Kimura's theory predicts that base substitutions should accumulate linearly with time and show little correlation with the magnitude of phenotypic changes. Ohta expanded the theory (in 1973 and 1996) to include not only a large number of neutral and of deleterious mutations, but also a large number of mutations that are slightly deleterious; that is, that have a very small selective effect, and therefore evolve neutrally in small populations, but are negatively selected in large populations[91,92]. Today the neutral and nearly neutral theory of molecular evolution is a widely accepted null hypothesis for nucleotide sequence evolution.

Importantly, the neutral theory provides a theoretical evolutionary model (also called the neutral model), which allows for the development of tests designed to find significant deviation from the substitution patterns predicted by the model. Such unusual substitution patterns probably indicate positive selection or influences other than negative selection and therefore suggest that a certain change or group of changes is of functional significance. In the decades following Kimura's initial hypothesis, several statistical tests to identify positive selection patterns were introduced and have proved instrumental in finding genomic patterns bearing signatures of functional adaptations (for examples see REFS 93–97).

This means that only a small fraction of all DNA sequence changes between humans and chimpanzees are relevant to the functional differences between the species, making the identification of non-neutral changes a daunting task. Few studies have been able to bridge the gap between DNA sequence differences and phenotype differences. Nevertheless, there are several cases where protein changes have been plausibly linked to the evolution of human-specific cognitive features. These examples are described in full elsewhere[17,18] and are not the subject of this review.

Human transcriptome evolution

Although changes to both the structure and regulation of proteins are ultimately caused by changes in DNA sequences, there is a fundamental difference between studying sequences and expression levels. Whereas the genome sequence of an individual need only be determined once, expression levels change over developmental stages, across different tissues, and in different environments. In fact, the expression of genes measured as mRNA levels is best seen as part of the phenotype of the organism. Expression levels of individual transcripts depend on those of other transcripts — they are genetically complex. Furthermore, measurement of mRNA expression levels is difficult because they are continuous variables that can fluctuate from zero to extremely high concentrations. Because of this and other technical problems, mRNA expression measurements include a large 'noise' component. In addition, cross-species studies present other challenges[19–21] (BOX 2). Nevertheless, in recent years, several studies have used microarrays to investigate expression differences among humans and other primates. Here we discuss these studies and the many challenges that lie ahead.

Gene expression and the neutral theory

The neutral theory of evolution (BOX 1) is the accepted null model for the evolution of DNA sequences. It postulates that the vast majority of nucleotide sequence differences observed between species do not affect function, and that many or most mutations in nucleotide sequence are deleterious and therefore subject to negative natural selection. Therefore they never (or only rarely) come to fixation. An alternative possibility, which might apply at the phenotypic level, is that most differences between species are adaptive and fixed by positive selection. To gauge which theory is a plausible model for evolution at the level of gene expression, it is useful to examine whether observations of expression evolution contradict any of the main predictions of the neutral theory. A central expectation is that the rate of expression evolution is mainly dependent on the mutation rate and on the amount of constraint acting on gene expression levels. Although both of these parameters are difficult to estimate (see below), two predictions arise. First, expression divergence between species should increase with evolutionary divergence, that is, with time. Second, the variation in gene expression among individuals of a species should be a function of the variation in gene expression among species. A brief review of the observations that are relevant to these issues follows (see also REFS 22,23).

Observations of a wide-spread influence of negative selection on expression evolution have sometimes been taken to mean that the neutral theory does not apply to gene expression evolution. However, a large and indeed dominant role of negative selection is an integral part of the neutral theory[16]. Negative selection has been demonstrated to have a strong influence on the rates at which gene expression changes both within and between species of fruitfly[24–26], and also in studies of mutation accumulation lines in flies and nematode worms[27,28]. Because flies and worms have much larger population sizes than primates, one might expect that negative selection would be a stronger factor in their evolution than for primates (BOX 1). However, negative selection has also been found to be a dominant factor in primates[26,29,30] and, despite previous observations[31,32], between humans and mice if one takes into account the effect of measuring orthologous genes with different arrays[33].

One observation that is in apparent disagreement with a dominant role of negative selection is that 23 expressed pseudogenes, which might lack a function and therefore evolved in the absence of both negative and positive selection, do not differ more in their expression between humans and chimpanzees than intact genes[19]. However, only a few pseudogenes were analysed, and it is also plausible that some of them do have functions (see for example REFS 34,35) or that *trans*-regulatory factors affecting functional target genes alter the expression of pseudogenes as a 'side effect'.

The neutral theory furthermore predicts that differences between species should accumulate approximately linearly with time (BOX 1). An increase in expression divergence with increasing genetic distance has been shown in fruitflies[24]. Similarly, the expression of most genes in different naturally occurring strains of yeast diverges proportionally to the genetic distance. Only a few genes in this study showed a correlation with phenotypic adaptations, presumably as a result of adaptations that are driven by positive selection[36]. In a recent study of natural populations of fish[37] the expression of 329 metabolic genes was analysed in the liver. About 18% of genes co-vary with the phylogenetic relationships between the fish populations, whereas only about 4% seem to be affected by positive selection when phylogenetic effects are taken into account. When expression differences are summarized across all genes with detectable levels of expression among humans, chimpanzees, orangutans and rhesus macaques, in the brain as well as the liver, the differences accumulate approximately linearly with time (as gauged from DNA sequence divergence). The same is true among three species of mice[19]. In apparent contrast to this, another study found that the expression patterns of 907 genes in the livers of humans, chimpanzees, orangutans and rhesus macaques seem to be equally distant from each other irrespective of their phylogenetic distance[30]. Although this observation requires further investigation, most data indicate that the extent of gene expression differences increases monotonically with evolutionary time, and that the rate with which this happens can be constant, at least for closely related species in which negative selection has not yet led to a saturation effect[22,38].

The neutral theory predicts that genes that vary less within a species will also tend to accumulate less change between species as these genes are subject to more negative selection and/or lower mutation rates. When

Box 2 | **Measuring gene expression in different primate species**

A number of experimental and technical issues make the study of gene expression in different species more challenging than in a single species.

First, at least for comparisons of humans and apes, it is not possible to obtain samples for which the environment has been controlled. In the future, it is likely that some environmental effects can be identified; for example, by manipulating the environment of experimental animals or by projects such as our proposed 'Human Transcriptome Project'. Of course, some environmental effects between species can be seen as *trans* effects of genetic factors that cause them. For example, dietary differences between species are dictated by genetic differences that govern how nutrients are processed in the digestive tract. Other differences include the effects of the niches that humans and chimpanzees carved for themselves during their evolution, such as the calming effects of grooming, or the use of medications.

Second, for ethical as well as practical reasons, tissues available for gene expression studies in humans and apes are almost always from cadavers, and are removed several hours after death. Of course, samples are collected from organs that are not affected by any known disease, or by conditions such as asphyxia or acidosis prior to death, which could affect gene expression patterns[98,99]. It is encouraging that in a study where gene expression in the human brain was compared between samples from autopsies and from surgical procedures, about 85% of gene expression differences between the hippocampus and the cerebral cortex seen in the living individuals were also found in the autopsy material[100].

Third, nucleotide sequence differences between species pose a problem because expression microarrays generally use probes to detect transcripts that are designed on the basis of the nucleotide sequences determined in only one species. This obviously results in problems when such an array is used in another species where consistent nucleotide differences might affect the efficiency with which the probes bind to transcripts. Sadly, it does not help to construct specific arrays for each species studied as expression levels cannot be compared across probes[33]. Earlier studies largely ignored this problem[101,102] or avoided it by limiting comparisons to between tissues within species[72,73]. Two approaches to deal with these problems have so far been used. For species where large amounts of genome or cDNA sequence are available, it is possible to identify oligonucleotides that are identical in sequence in the two species[19,20,29,78]. This approach works only for fairly closely related species because the number of probes retained drops quickly with the sequence divergence between the species. However, even identical probes could cause apparent consistent expression differences due to sequence difference in the transcript outside the target region if they affect the secondary structure of the transcripts. This can also occur when probes designed to fit a transcript in one species fall outside a transcript in another species owing to differences in polyadenylation, splicing or other effects. Cross-hybridization of probes to other transcripts that might differ in their expression and sequence between the species can also cause problems. To overcome them it is helpful to use other filters to detect probes that yield signals that are inconsistent with other probes that detect the same transcript[40].

A second useful approach uses arrays that contain PCR-amplified cDNA probes for all species analysed[21,30]. This has the advantage that cDNA arrays are considered less susceptible to problems caused by sequence differences outside the target region. However, their technical noise level is generally larger than standardized oligonucleotide arrays. It is also difficult, perhaps impossible, to successfully amplify a large proportion of all putative transcripts from many species, and limiting the arrays to only successful amplification products could result in a biased selection of transcripts.

Eventually, hybridization-based techniques are likely to become obsolete as they have the drawback that the absolute number of transcript molecules present in samples cannot be determined. Methods that measure absolute numbers of molecules such as high-throughput parallel sequencing[103] of cDNAs are likely to become the technique of choice.

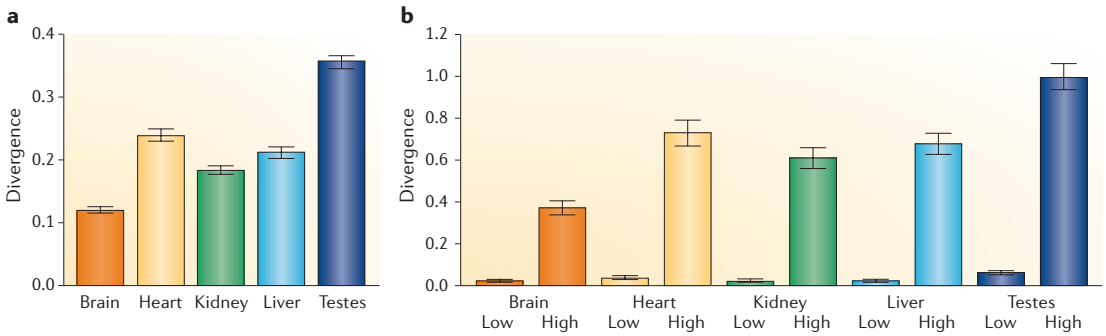

Figure 1 | **Expression divergence between humans and chimpanzees in different tissues.** Bar heights represent expression divergence between mean expression levels in 6 humans and 5 chimpanzees for all expressed genes (**a**), and for genes with the 10% lowest (low) and 10% highest (high) expression variation among the 6 humans (**b**) (data from REF. 29). Colours represent different tissues: orange represents brain, yellow represents heart, green represents kidney, light blue represents liver and dark blue represents testes. Error bars show 95% confidence intervals calculated by 1,000 bootstraps over all genes expressed in a tissue. As can be seen, the brain shows the least divergence between humans and chimpanzees, indicating that negative selection has the biggest effect on this tissue. Furthermore, genes showing low (or high) variation within humans, show low (or high) divergence between humans and chimpanzees, also indicating that constraints have a large effect on expression evolution.

907 genes expressed in heart muscle were analysed in fish on the Atlantic coast of North America, most of the expression variation among populations was found to reflect variation within populations[39]. In humans, genes expressed in the brain that vary more among individuals tend to differ more in their expression between humans and chimpanzees (FIG. 1), as well as between humans and other primates[19]. The same holds true for genes expressed in the brains of three species of mice[19].

A further prediction of the neutral theory is that the extent of overall molecular change is independent of the extent of phenotypic change. When humans and chimpanzees are compared, gene expression differs much less between species in the brain than in the liver, kidneys, heart or testes[29] (FIG. 1), contrary to an intuitive expectation that the function of the brain has changed more between humans and apes than the function of the other tissues. Even more counterintuitive is that gene expression in regions of the cerebral cortex that are involved in human-specific functions such as speech production do not differ more between the two species than others that do not seem to have changed their function, such as the primary visual cortex[40]. However, this is not unexpected if functional constraint and therefore negative selection is the primary force determining expression change, as predicted by the neutral theory. Genes expressed in the brain might be involved in more functional interactions than genes expressed in the other organs and would therefore be under more functional constraints. In fact, it has been shown in yeast that the expression of genes that are involved in more interactions evolves slower than the expression of genes that are involved in fewer interactions[41].

Of special interest is the comparison of the rates and mode of evolution of mRNA expression levels and of the sequences of the proteins encoded by these mRNAs. If the neutral theory applies at both levels, one might expect to find parallel patterns of change. Among five tissues in chimpanzees and humans, the brain has the lowest divergence at both levels, the heart and kidneys have diverged more, whereas at both levels the liver and testes are the most diverged[29]. When expression in entire organisms of two *Drosophila* species are compared, a correlation between amino-acid differences and expression differences similarly exists across genes[25,42]. Nuzhdin *et al.*[25] have used this observation to argue for wide-spread positive selection. Because amino-acid changes have been inferred to be largely under positive selection in flies[43,44], these authors concluded that the parallel patterns of variation in expression and sequence imply that a substantial fraction of expression divergence is adaptive in flies. However, such a large extent of positive selection on amino-acid differences has not been seen in humans[14], possibly owing to their smaller effective population size (BOX 1). A hint at the relative extent of positive selection, negative selection and neutral fixation comes from genes that are expressed in multiple tissues in primates. If positive selection is the main force, the more tissues a gene is expressed in, the more divergent selection pressures it will be exposed to, and therefore the more it will have changed in different tissues. In contrast to this, it is observed that the more tissues a gene is expressed in the less it changes between species, and that these expression changes tend to be similar in the different tissues. So among humans and apes it seems that functional constraints that are similar in their relative magnitude at the levels of mRNA expression and amino-acid sequence limit the divergence between primate species and that these constraints are additive across tissues (FIG. 2).

In conclusion, most studies of gene expression evolution to date are compatible with the notion that most expression differences between species are selectively neutral or nearly neutral. This does not necessarily mean that the neutral theory represents an ultimately correct description of the mode of transcriptome

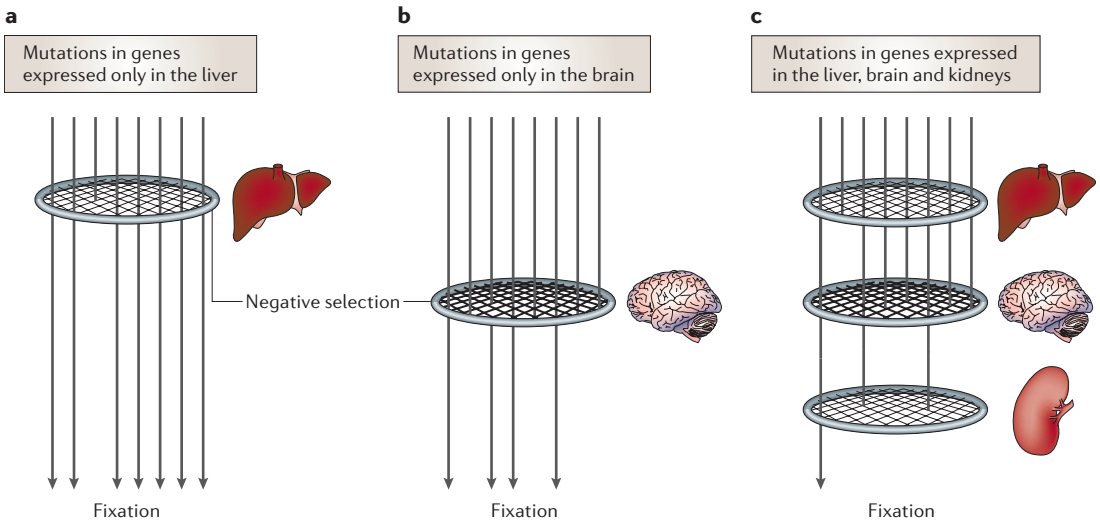

Figure 2 | **Negative selection adds up across tissues.** Each panel shows several mutations that affect expression levels of genes only in the liver (**a**), only in the brain (**b**), or in the brain, liver and kidneys (**c**). As there are more constraints acting in the brain than in the liver, more mutations are weeded out by negative selection in the brain than in the liver. For genes expressed in several tissues (**c**) a mutation needs only to be detrimental in one tissue to be weeded out by negative selection. Therefore, even more mutations are weeded out by negative selection, leading to the tendency for genes that are expressed in more tissues to be less diverged between species[29]. The same scenario would also apply for mutations that affect the protein sequence of genes[105].

evolution. Current observations can also be explained by pervasive positive selection and other factors, but such scenarios are more complex and require different modes of selection for different observations. As such, the neutral model is the most parsimonious explanation of the current data, although comparisons among more species will be necessary to establish its validity further. The neutral theory has the extra advantage that it is a null hypothesis against which observations can be tested to identify groups of genes or organs in which its predictions are rejected. To date, there are two main areas in which signs for positive selection on gene expression in primates are emerging, which we discuss below.

Positive selection in testes

The neutral theory predicts that divergence between species should be positively correlated with diversity within species. If positive selection occurs, the ratio of divergence between species to diversity within species is expected to increase[45]. Several groups have suggested that this expectation could be used to detect positive selection on gene expression[19,24,25]. However, this ratio depends on many parameters, such as environmental influence, demography, mutation rate, divergence time and selective constraint. If one compares divergence with diversity ratios for genes or sets of genes, differences in these parameters need to be considered. When expression divergence to diversity ratios are compared across tissues between humans and chimpanzees, they are about equal in the brain, heart, kidneys and liver, but roughly threefold higher in the testes[29]. A possible explanation for this relatively low diversity is that, although genes expressed in the testes are under little constraint (which would normally lead to a large diversity), expression in this tissue is much less influenced by the environment than expression in other tissues. This would lead to a lower observed diversity in testes than in other tissues. But if there is little constraint on the genes expressed in the testes, we would expect that the testes also impose little constraint on the expression levels of these same genes in other tissues. By contrast, however, the expression levels in somatic tissues of genes that are also expressed in the testes are more constrained than those of genes that are not expressed in the testes[29].

A further indication that positively selected gene expression changes are frequent in the testes is that an excess of the genes that differ in expression in the testes between humans and chimpanzees are located on the X chromosome[29]. This would be expected for positively selected changes that are advantageous in males and genetically recessive, as males carry only one X chromosome and so experience an immediate phenotypic effect of such mutations[46]. There is also an excess of amino-acid changes in proteins expressed in the testes and encoded on the X chromosome[29]. More work on modelling gene expression evolution (see below) is needed to gauge the extent of positive selection in testes, and comparisons in more tissues and among more primate species are needed to explore how outstanding gene expression evolution in the testes in humans and chimpanzees is. Nevertheless, current observations justify the hypothesis that positive selection shapes gene expression patterns considerably more often in the testes than in other organs.

Several other studies have shown similar indications for positive selection at the amino-acid sequence level of genes expressed in the testes. For example, in

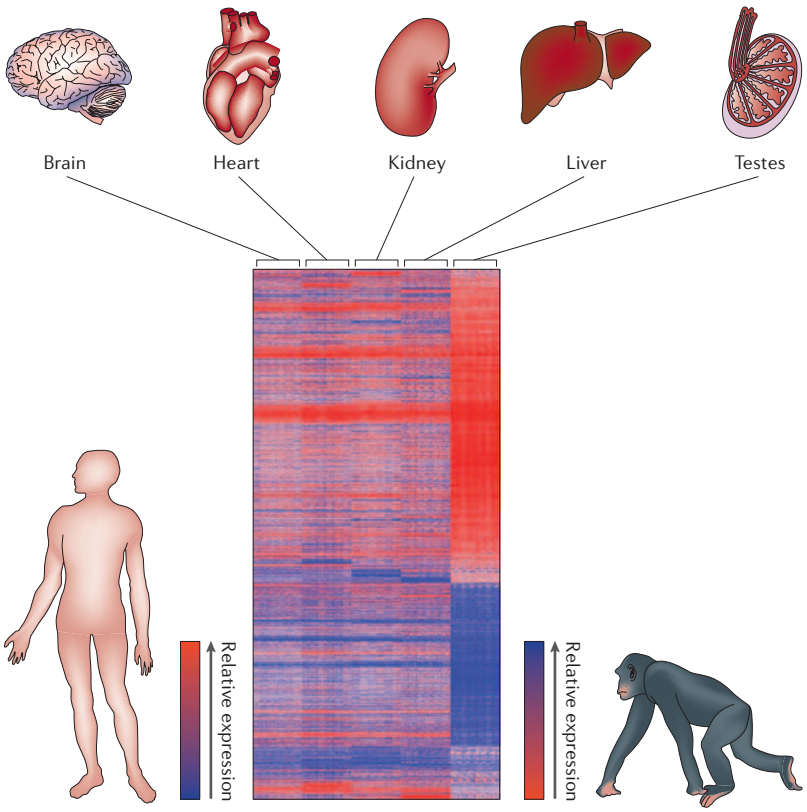

Figure 3 | **Hierarchical clustering of expression differences between humans and chimpanzees in five different tissues.** All probe sets differently expressed between humans and chimpanzees in at least one tissue are shown (data from REF. 29). Genes in red are more highly expressed in humans than in chimpanzees and genes in blue represent the reverse. Note that the testes exhibit many more differences than the other four tissues. Expression profile reproduced from REF. 29 © (2005) American Association for the Advancement of Science.

a genome-wide comparison of gene sequences from humans and chimpanzees, genes that are most highly expressed in the testes (of 28 tissues analysed) were most likely to be positively selected at the amino-acid level[47]. Furthermore, genes involved in spermatogenesis are overrepresented among positively selected genes in primates[47,48] — and potentially also among those that have been selected between human groups[49]— as are genes encoding proteins present in the seminal fluid[50]. These evolutionary patterns seem to be common to the human and the chimpanzee lineages both with respect to gene expression and amino-acid sequence, indicating that positive selection influenced the male reproductive system during the evolution of both species. In fact, positive selection for traits related to reproduction, and male-specific reproduction in particular, could be a general phenomena in a wide variety of species[51,52]. The evidence is especially compelling in *Drosophila* where — similar to what is found between humans and chimpanzees — recent studies have shown that genes with a male-biased expression evolve particularly fast in terms of protein sequence and gene expression[53–56]. Interestingly, however, genes involved in spermatogenesis might have evolved even faster during primate evolution than during rodent or dog evolution[14,57].

If the fraction of positively selected gene expression changes is substantially higher in the testes than in most other tissues, it implies that many genes are regulated differently in the testes than elsewhere in the body, as mutations that are beneficial for male reproduction might be detrimental in other tissues (for examples see REF. 58). Indeed, it is known that transcriptional regulation, alternative splicing, transcription start-site usage, translational regulation and several other aspects of gene expression differ between cells involved in spermatogenesis and those in somatic tissues[59,60]. This is supported by the observation that gene expression in the testes shows more independent changes between species than in other tissues[29](FIG. 3). In the future, it will be important to compare various aspects of gene expression in the testes to other tissues in a systematic and genome-wide way.

What are the selective pressures that might lead to more positive selection in the testes? Sperm competition between males is thought to be a primary force for the rapid evolution of male reproductive traits in general, and the large differences in testis size among humans and chimpanzees in particular[61]. Together with sexual conflict, it is also thought to cause the rapid evolution of genes involved in traits such as sperm–egg recognition or seminal fluid composition[51]. However, it is not immediately obvious why sperm competition should have such a wide-spread effect on such a variety of genes expressed in the testes, including genes encoding proteins that are involved in very basic functions, such as protamin 1 and 2 (REFS 47,48), especially as multiple female copulations are not a universal trait among apes[62]. Comparisons among more primate species will clarify how well changes in competition between the sexes, or in mating systems that would affect sperm competition, correlate with gene expression changes in the testes.

An alternative and not mutually exclusive factor that might exert positive selection on gene expression in the testes is meiotic drive, which would occur if the fitness of spermatids or spermatozoa were controlled by their haploid genomes. This could potentially shape and change fitness landscapes of genes expressed in the testes. However, although meiotic drive can occur in mammals[63], there are some indications that sperm phenotype is controlled by the diploid genome[64–66] (for example, through cytoplasmic bridges between spermatids[67,68]) and that more mature spermatids do not transcribe genes[68]. Studies of gene expression in isolated single spermatids would be a first step towards estimating the potential for meiotic drive in the male germ line of primates.

Irrespective of whether sperm competition, meiotic drive or some other mechanism is at play, male reproduction in particular is an area where humans and chimpanzees have accumulated many molecular differences. In fact, if a certain gene has been positively selected during primate evolution and is expressed in the testes as well as other tissues, it might be more likely that the cause of selection has to do with male reproduction than with other functions. This is especially true if the evidence for selection is not restricted to a single evolutionary lineage and if immunity-related

Sperm competition
The direct competition between sperm from different males that occurs when females copulate with multiple males.

Meiotic drive
Preferential transmission of one of two alleles from a parent to its offspring.

functions seem unlikely. In such cases (for examples see REFS 69–71), we feel that speculations and preliminary investigations that are aimed at finding the cause of positive selection should focus on the male germ line.

Putative positive selection in the human brain

Gene expression in the human brain has attracted particular interest because many of the phenotypic differences that set humans apart from other organisms have a cognitive component. One might therefore expect the human brain to have diverged more from the brains of other animals than other organs have diverged from their homologues. Contrary to this expectation, however, gene expression in the brain has diverged less than other organs analysed to date (FIG. 1). Furthermore, when various regions of the brain are analysed, parts of the brain that are involved in human-specific traits such as the production of language — for example, Broca's area — have not diverged more in their gene expression than regions such as the primary visual cortex, which as far as is known has not changed its function[40]. This is in agreement with the neutral theory that postulates that, overall, the extent of difference should not reflect the extent of functional difference but rather the evolutionary time since two organisms or two tissues diverged (BOX 3).

There is, however, a subtle indication that positive selection could have affected gene expression in the brain during recent human evolution. This comes from the observation that when an outgroup is used to assign gene expression changes since the common ancestor to either the human or chimpanzee lineages, the ratio of change on the human to the chimpanzee lineage is larger in the brain than in the liver or heart[72–75]. This finding is also seen when expression changes are assigned to lineages on the basis that the magnitude of expression changes tend to be asymmetrical such that upregulations are fewer but larger in magnitude. The species with the skew towards upregulated expression changes is assumed to be the one that has changed more since the common ancestor[29,76]. So of the relatively few expression changes that have occurred in the brain, relatively more seem to have happened on the human than the chimpanzee lineage. It should be noted that this does not mean that the expression of a large proportion of genes must have changed more in the human than the chimpanzee brain. The observed effect can easily be explained by as few as 10% of all genes[76]. Therefore, the accelerated evolution of brain gene expression on the human lineage is a subtle effect that requires further investigation. Interestingly, there is some evidence that genes expressed in the brain might also have changed more in their amino-acid sequences on the human than on the chimpanzee lineage[29,77], although this effect might be slight[29].

If an acceleration of the evolution of brain gene expression exists, it could either be due to more genes being positively selected on the human lineage, or a relaxation of constraints and therefore less negative selection on genes expressed in the human brain than in the chimpanzee brain. The second suggestion might seem counterintuitive but is a formal possibility that has to be considered. However, two lines of evidence suggest that it is positive selection that has affected genes expressed in the human brain. The first observation is that the vast majority of regulatory changes on the human lineage have led to an increase in expression level[73–75]. This is not easily compatible with a model in which the intensity of negative selection has been reduced in humans as that should not lead to a directional change of expression levels. It remains possible, however, that this effect is due to upregulations being larger in magnitude than downregulations, and therefore more readily detectable[76]. Future work should clarify this.

The second observation is that when gene expression changes in the brain, heart, kidneys and liver of humans and chimpanzees are assigned to the lineages using outgroup species, it is only in the brain that groups of functionally related genes that changed their expression more on the human lineage are associated with larger areas of linkage disequilibrium in three human populations; Africans, Chinese and Europeans[78]. Increased levels of linkage disequilibrium are an indirect indicator of positive selection. As alternative explanations would need to account for the fact that this effect is seen in brain but not in three other tissues, positive selection seems likely.

It is tantalizing that linkage disequilibrium is a transitory sign of positive selection and is unlikely to be detected if the selective events are more than 200,000 years old[79]. It is also interesting that the groups of genes in this analysis that have the highest signals of selection and of human expression change tend to be involved in energy production and are more highly expressed in human brains than in chimpanzee brains. Therefore, at least some of the events that affected the human brain could be recent, and could involve upregulations of energy metabolism.

Future directions

It is obvious that our understanding of the evolution of gene expression is rudimentary compared with our understanding of the evolution of gene sequences. One reason for this is that the expression of a gene is a dynamic and continuous variable that changes with developmental and physiological states. Current technologies can capture this only in a very crude way at best, especially in humans and apes where one is largely limited to the study of individuals that have died of natural causes. Another difficulty is that current methods of analysing organs and tissues rely largely on samples that represent an average of millions of cells of several types, removed from an organ of a dead individual. Hopefully, novel technologies will address these issues. For example, laser capture microdissection of small numbers of cells of one type can be used for analyses of gene expression[80,81]. Such techniques should soon be developed for application to needle biopsies from healthy living individuals. This would greatly advance our understanding of gene expression in health, disease and evolution.

Outgroups
Species that are more distantly related to two or more species studied and can therefore be used to estimate the ancestral state of a trait such as nucleotide sequence or gene expression level.

Linkage disequilibrium
Non-random association of nucleotide polymorphisms along the chromosome. Larger areas of stronger linkage disequilibrium are seen around a genetic change that has been positively selected in the recent past.

Another fundamental limitation to our current understanding of gene expression evolution, and therefore our ability to identify positive selection in the evolutionary past, is the absence of an adequate model of gene expression evolution. Several studies have made a start at this[24,32,37,75,76] and, as discussed above, the neutral theory is, in our view, a logical starting place. However, many important parameters that are necessary to build a realistic model remain to be elucidated. They include basic parameters such as estimates of the relative proportions of *cis*- and *trans*-regulatory changes, and of the rate and mode of the changes in expression that underlie the differences observed within and between species. The extent to which regulatory changes 'spill over' to cell types and tissues where they have no functional role and where the transcript might not even be translated is another unknown factor[82]. Much greater mechanistic knowledge of gene expression evolution is required to develop realistic evolutionary models that allow the identification of adaptive regulatory changes. This remains an immense challenge, but approaches such as the mapping of expression differences by linkage analysis in pedigrees of humans and laboratory animals (for examples see REF. 83) promise progress.

Box 3 | **The evolution of distinct tissue expression patterns**

In addition to understanding the patterns of gene expression changes within tissues, it is instructive to understand how differences between tissues evolve over time. Two studies looked at gene expression in a large set of tissues in humans and mice[84,85]. Only very few genes were found to be expressed exclusively in a single tissue — with the testes having most tissue-specific genes. At the other extreme, the studies found that only 3–6% of genes were expressed in all tissues. Using species comparisons, we can try to understand how these differences in expression have evolved over time. Khaitovich *et al.* observed a very slow accumulation of differences between tissues over time[40]. Indeed, distinct cortical regions involved in different cognitive functions, such as Broca's area, its mirror area from the right hemisphere, the dorsolateral prefrontal cortex, and the pre-motor cortex have statistically indistinguishable expression patterns in the human brain (P.K., unpublished observations). Correspondingly, all these brain regions show the same expression differences between humans and chimpanzees.

We currently have little understanding of how expression differences between two tissues are created, but variation in tissue-specific expression patterns within a species can be substantial[104]. Both tissues always share the same genotype, and therefore the process seems to be different from the process that creates expression differences between two species — each of which has a different genome. However, we can ask whether most of the differences that accumulate between tissues fix in a species through random drift, or whether most of them fix owing to positive selection. Again, there will be mutations that will not be observed, because they are weeded out by negative selection. Within human, chimpanzee and mouse brains, differences between various brain regions accumulate proportionally to the time since the last common ancestor in which the regions were not diverged[19]. This could be because differences between tissues accumulate through drift, and are monotonic with time for that reason. Or it could be that differences accumulate through a selection process on function, but that this process sometimes creates a monotonic increase with time, and that this was the case in the brain. It could also be that differences between tissues are monotonic in ontological time, but that that process is often aligned with, or recapitulates, evolutionary time. Yanai *et. al* explored different explanations for these phenomena[82]. Further studies will need to resolve these possibilities. Irrespective of which of these possibilities hold true, if differences in expression between tissues increase with time since the last common ancestor in which the tissues had not diverged, then large-scale expression measurements can provide a new tool to reconstruct the evolutionary history of the differentiation of tissues.

A Human Transcriptome Project?

In spite of the shortcomings inherent to current technologies, reliable estimates of gene expression for all human transcripts, including independent measurements of transcript isoforms, can be obtained using established microarray technologies. However, although expression data collected from many different human and mouse tissues from a limited number of individuals[84,85] have yielded interesting results (BOX 3), no organized effort exists for the collection of human expression phenotypes from many individuals. Because the phenotypic changes observed in complex human disorders and normal human phenotypes are likely to be caused by an interplay of both structural and regulatory factors, a combined analysis of gene expression and genome variation within and between species would be immensely valuable. The genome sequencing projects and the International HapMap Project[86] that studies genomic variation among humans have shown how successful large collaborative projects using standardized technology platforms can be. An organized effort to study gene expression levels in large numbers of human individuals, in conjunction with genome variation in the same individuals, would substantially enhance the value of the HapMap project. Therefore, we call for a 'Human Transcriptome Project' that is dedicated to the collection of gene expression data in combination with nucleotide variation among humans as well as in a few related primate species where genome sequences have been generated.

Conclusions

Ever since it was first suggested more than 30 years ago[1], the question of whether regulatory protein changes are more important than structural ones for the human phenotype has often been discussed. In light of the parallel patterns of evolution of protein sequences and gene expression levels[29], this question seems to be a fruitless one. When there is positive selection for novel functions in an organ system, both structural changes and expression changes will occur to bring about a change in phenotype. This is intuitive; for example, increased enzymatic activity in a tissue can occur through increased expression of the relevant enzyme or by structural changes that increase enzyme activity. Often, both things will happen. It is therefore likely that both regulatory and structural changes are responsible for the emergence of the human phenotype. It should be noted that the strong genome-wide effects of positive selection that have been detected in primates — for example, for genes expressed in the testes or in the immune system — are not specific to the human lineage, but rather common to primates or probably even mammals, and are therefore informative about general evolutionary patterns. The only genome-wide feature specific to humans so far detected is the acceleration of evolution of genes expressed in the brain, but this is a relatively weak effect. To identify the specific regulatory or structural changes responsible for human-specific features, genome-wide surveys of gene expression and gene sequences are exciting and necessary starting points that, when successful, will generate hypotheses that must then be tested in cell culture or animal models.

1. King, M. C. & Wilson, A. C. Evolution at two levels in humans and chimpanzees. *Science* **188**, 107–116 (1975).
2. Szathmary, E. & Smith, J. M. The major evolutionary transitions. *Nature* **374**, 227–232 (1995).
3. Breuer, T., Ndoundou-Hockemba, M. & Fishlock, V. First observation of tool use in wild gorillas. *PLoS Biol.* **3**, e380 (2005).
4. Galdikas, B. M. Orangutan tool use. *Science* **243**, 152 (1989).
5. Boesch, C. & Boesch, H. Tool use and tool making in wild chimpanzees. *Folia Primatol. (Basel)* **54**, 86–99 (1990).
6. Sugiyama, Y. Tool use by wild chimpanzees. *Nature* **367**, 327 (1994).
7. Phillips, K. A. Tool use in wild capuchin monkeys (*Cebus albifrons trinitatis*). *Am. J. Primatol.* **46**, 259–261 (1998).
8. Ducoing, A. M. & Thierry, B. Tool-use learning in Tonkean macaques (*Macaca tonkeana*). *Anim. Cogn.* **8**, 103–113 (2005).
9. Whiten, A. *et al.* Cultures in chimpanzees. *Nature* **399**, 682–685 (1999).
10. van Schaik, C. P. *et al.* Orangutan cultures and the evolution of material culture. *Science* **299**, 102–105 (2003).
11. Avital, E. & Jablonka, E. *Animal Traditions: Behavioural Inheritance in Evolution* (Cambridge Univ. Press, Cambridge; New York, 2000).
12. Whiten, A. The second inheritance system of chimpanzees and humans. *Nature* **437**, 52–55 (2005).
13. Glazko, G. V. & Nei, M. Estimation of divergence times for major lineages of primate species. *Mol. Biol. Evol.* **20**, 424–434 (2003).
14. Mikkelsen, T. *et al.* Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature* **437**, 69–87 (2005).
This paper discusses human and chimpanzee evolution on the level of genomic DNA sequence.
15. Wienberg, J. Fluorescence *in situ* hybridization to chromosomes as a tool to understand human and primate genome evolution. *Cytogenet. Genome Res.* **108**, 139–160 (2005).
16. Kimura, M. Evolutionary rate at the molecular level. *Nature* **217**, 624–626 (1968).
17. Varki, A. & Altheide, T. K. Comparing the human and chimpanzee genomes: searching for needles in a haystack. *Genome Res.* **15**, 1746–1758 (2005).
18. Gilbert, S. L., Dobyns, W. B. & Lahn, B. T. Genetic links between brain development and brain evolution. *Nature Rev. Genet.* **6**, 581–590 (2005).
19. Khaitovich, P. *et al.* A neutral model of transcriptome evolution. *PLoS Biol.* **2**, e132 (2004).
This paper summarizes how the neutral theory can be applied to gene expression evolution.
20. Preuss, T. M., Caceres, M., Oldham, M. C. & Geschwind, D. H. Human brain evolution: insights from microarrays. *Nature Rev. Genet.* **5**, 850–860 (2004).
21. Gilad, Y., Rifkin, S. A., Bertone, P., Gerstein, M. & White, K. P. Multi-species microarrays reveal the effect of sequence divergence on gene expression profiles. *Genome Res.* **15**, 674–680 (2005).
22. Whitehead, A. & Crawford, D. L. Variation within and among species in gene expression: raw material for evolution. *Mol. Ecol.* **15**, 1197–1211 (2006).
23. Ranz, J. M. & Machado, C. A. Uncovering evolutionary patterns of gene expression using microarrays. *Trends Ecol. Evol.* **21**, 29–37 (2006).
24. Rifkin, S. A., Kim, J. & White, K. P. Evolution of gene expression in the *Drosophila melanogaster* subgroup. *Nature Genet.* **33**, 138–144 (2003).
This paper describes the expression evolution in *Drosophila melanogaster* and demonstrates that expression differences correlate with divergence time as measured from DNA sequences.
25. Nuzhdin, S. V., Wayne, M. L., Harmon, K. L. & McIntyre, L. M. Common pattern of evolution of gene expression level and protein sequence in *Drosophila*. *Mol. Biol. Evol.* **21**, 1308–1317 (2004).
This paper argues that positive selection is a substantial factor in expression evolution in *Drosophila*.
26. Lemos, B., Meiklejohn, C. D., Caceres, M. & Hartl, D. L. Rates of divergence in gene expression profiles of primates, mice, and flies: stabilizing selection and variability among functional categories. *Evolution Int. J. Org. Evolution* **59**, 126–137 (2005).
27. Denver, D. R. *et al.* The transcriptional consequences of mutation and natural selection in *Caenorhabditis elegans*. *Nature Genet.* **37**, 544–548 (2005).
This paper shows that negative selection has a dominant role in gene expression evolution in *Caenorhabditis elegans*.
28. Rifkin, S. A., Houle, D., Kim, J. & White, K. P. A mutation accumulation assay reveals a broad capacity for rapid evolution of gene expression. *Nature* **438**, 220–223 (2005).
This paper shows that negative selection has a dominant role in gene expression evolution in *Drosophila*.
29. Khaitovich, P. *et al.* Parallel patterns of evolution in the genomes and transcriptomes of humans and chimpanzees. *Science* **309**, 1850–1854 (2005).
This paper demonstrates that modes of protein sequences and gene expression evolution are similar to one another, but differ among tissues in primates.
30. Gilad, Y., Oshlack, A., Smyth, G. K., Speed, T. P. & White, K. P. Expression profiling in primates reveals a rapid evolution of human transcription factors. *Nature* **440**, 242–245 (2006).
31. Jordan, I. K., Marino-Ramirez, L., Wolf, Y. I. & Koonin, E. V. Conservation and coevolution in the scale-free human gene coexpression network. *Mol. Biol. Evol.* **21**, 2058–2070 (2004).
32. Yanai, I., Graur, D. & Ophir, R. Incongruent expression profiles between human and mouse orthologous genes suggest widespread neutral evolution of transcription control. *OMICS* **8**, 15–24 (2004).
33. Liao, B. Y. & Zhang, J. Evolutionary conservation of expression profiles between human and mouse orthologous genes. *Mol. Biol. Evol.* **23**, 530–540 (2006).
34. Korneev, S. A., Park, J. H. & O'Shea, M. Neuronal expression of neural nitric oxide synthase (nNOS) protein is suppressed by an antisense RNA transcribed from an NOS pseudogene. *J. Neurosci.* **19**, 7711–7720 (1999).
35. Hirotsune, S. *et al.* An expressed pseudogene regulates the messenger-RNA stability of its homologous coding gene. *Nature* **423**, 91–96 (2003).
36. Fay, J. C., McCullough, H. L., Sniegowski, P. D. & Eisen, M. B. Population genetic variation in gene expression is associated with phenotypic variation in *Saccharomyces cerevisiae*. *Genome Biol.* **5**, R26 (2004).
37. Whitehead, A. & Crawford, D. L. Neutral and adaptive variation in gene expression. *Proc. Natl Acad. Sci. USA* ***103***, 5425–5430 (2006).
This paper uses expression variation within and between populations to assess influence of positive and negative selection on expression evolution in teleost fish.
38. Jordan, I. K., Marino-Ramirez, L. & Koonin, E. V. Evolutionary significance of gene expression divergence. *Gene* **345**, 119–126 (2005).
39. Oleksiak, M. F., Churchill, G. A. & Crawford, D. L. Variation in gene expression within and among natural populations. *Nature Genet.* **32**, 261–266 (2002).
40. Khaitovich, P. *et al.* Regional patterns of gene expression in human and chimpanzee brains. *Genome Res.* **14**, 1462–1473 (2004).
41. Fraser, H. B., Hirsh, A. E., Steinmetz, L. M., Sharfe, C. & Feldman, M. W. Evolutionary rate in the protein interaction network. *Science* **296**, 750–752 (2002).
42. Lemos, B., Bettencourt, B. R., Meiklejohn, C. D. & Hartl, D. L. Evolution of proteins and gene expression levels are coupled in *Drosophila* and are independently associated with mRNA abundance, protein length, and number of protein–protein interactions. *Mol. Biol. Evol.* **22**, 1345–1354 (2005).
This paper demonstrates that modes of protein sequences and gene expression evolution are similar in *Drosophila*.
43. Smith, N. G. & Eyre-Walker, A. Adaptive protein evolution in *Drosophila*. *Nature* **415**, 1022–1024 (2002).
44. Fay, J. C., Wyckoff, G. J. & Wu, C. I. Testing the neutral theory of molecular evolution with genomic data from *Drosophila*. *Nature* **415**, 1024–1026 (2002).
45. Lynch, M. & Hill, W. G. Phenotypic evolution by neutral mutation. *Evolution* **40**, 915–935 (1986).
46. Rice, W. R. Sex chromosomes and the evolution of sexual dimorphism. *Evolution* **38**, 735–742 (1984).
47. Nielsen, R. *et al.* A scan for positively selected genes in the genomes of humans and chimpanzees. *PLoS Biol.* **3**, e170 (2005).
48. Wyckoff, G. J., Wang, W. & Wu, C. I. Rapid evolution of male reproductive genes in the descent of man. *Nature* **403**, 304–309 (2000).
49. Voight, B. F., Kudaravalli, S., Wen, X. & Pritchard, J. K. A map of recent positive selection in the human genome. *PLoS Biol.* **4**, e72 (2006).
50. Clark, N. L. & Swanson, W. J. Pervasive adaptive evolution in primate seminal proteins. *PLoS Genet.* **1**, e35 (2005).
51. Swanson, W. J. & Vacquier, V. D. The rapid evolution of reproductive proteins. *Nature Rev. Genet.* **3**, 137–144 (2002).
52. Good, J. M. & Nachman, M. W. Rates of protein evolution are positively correlated with developmental timing of expression during mouse spermatogenesis. *Mol. Biol. Evol.* **22**, 1044–1052 (2005).
53. Meiklejohn, C. D., Parsch, J., Ranz, J. M. & Hartl, D. L. Rapid evolution of male-biased gene expression in Drosophila. *Proc. Natl Acad. Sci. USA* **100**, 9894–9899 (2003).
54. Ranz, J. M., Castillo-Davis, C. I., Meiklejohn, C. D. & Hartl, D. L. Sex-dependent gene expression and evolution of the *Drosophila* transcriptome. *Science* **300**, 1742–1745 (2003).
55. Zhang, Z. & Parsch, J. Positive correlation between evolutionary rate and recombination rate in *Drosophila* genes with male-biased expression. *Mol. Biol. Evol.* **22**, 1945–1947 (2005).
56. Zhang, Z., Hambuch, T. M. & Parsch, J. Molecular evolution of sex-biased genes in *Drosophila*. *Mol. Biol. Evol.* **21**, 2130–2139 (2004).
57. Lindblad-Toh, K. *et al.* Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature* **438**, 803–819 (2005).
58. Goriely, A. *et al.* Gain-of-function amino acid substitutions drive positive selection of *FGFR2* mutations in human spermatogonia. *Proc. Natl Acad. Sci. USA* **102**, 6051–6056 (2005).
59. Kleene, K. C. A possible meiotic function of the peculiar patterns of gene expression in mammalian spermatogenic cells. *Mech. Dev.* **106**, 3–23 (2001).
60. Kleene, K. C. Patterns, mechanisms, and functions of translation regulation in mammalian spermatogenic cells. *Cytogenet. Genome Res.* **103**, 217–224 (2003).
61. Birkhead, T. R. & Pizzari, T. Postcopulatory sexual selection. *Nature Rev. Genet.* **3**, 262–273 (2002).
62. Robbins, M. M. *et al.* Social structure and life-history patterns in western gorillas (*Gorilla gorilla gorilla*). *Am. J. Primatol.* **64**, 145–159 (2004).
63. Lyttle, T. W. Cheaters sometimes prosper: distortion of mendelian segregation by meiotic drive. *Trends Genet.* **9**, 205–210 (1993).
64. Parker, G. A. & Begon, M. E. Sperm competition games: sperm size and number under gametic control. *Proc. Biol. Sci.* **253**, 255–262 (1993).
65. Parker, G. A. Sperm competition games: sperm size and sperm number under adult control. *Proc. Biol. Sci.* **253**, 245–254 (1993).
66. Haig, D. & Bergstrom, C. T. Multiple mating, sperm competition and meiotic drive. *J. Evol. Biol.* **8**, 265–282 (1995).
67. Handel, M. *Spermatogenesis: Genetic Aspects* Vol. 15 (ed. Hennig, W.) (Springer, Berlin, 1987).
68. Kierszenbaum, A. & Tres, L. L. Structural and transcriptional features of mouse spermatid genome. *J. Cell Biol.* **65**, 258–270 (1975).
69. Kouprina, N. *et al.* Accelerated evolution of the *ASPM* gene controlling brain size begins prior to human brain expansion. *PLoS Biol.* **2**, e126 (2004).
70. Evans, P. D. *et al.* Microcephalin, a gene regulating brain size, continues to evolve adaptively in humans. *Science* **309**, 1717–1720 (2005).
71. Kouprina, N. *et al.* The microcephaly *ASPM* gene is expressed in proliferating tissues and encodes for a mitotic spindle protein. *Hum. Mol. Genet.* **14**, 2155–2165 (2005).
72. Enard, W. *et al.* Intra- and interspecific variation in primate gene expression patterns. *Science* **296**, 340–343 (2002).
The first paper that compared gene expression patterns in primates and described a brain-specific acceleration in humans.
73. Caceres, M. *et al.* Elevated gene expression levels distinguish human from non-human primate brains. *Proc. Natl Acad. Sci. USA* **100**, 13030–13035 (2003).
74. Gu, J. & Gu, X. Induced gene expression in human brain after the split from chimpanzee. *Trends Genet.* **19**, 63–65 (2003).
75. 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).

76. Khaitovich, P., Paabo, S. & Weiss, G. Toward a neutral evolutionary model of gene expression. *Genetics* **170**, 929–939 (2005).
77. Dorus, S. *et al.* Accelerated evolution of nervous system genes in the origin of *Homo sapiens*. *Cell* **119**, 1027–1040 (2004).
78. Khaitovich, P. *et al.* Positive selection on gene expression in the human brain. *Curr. Biol.* **16**, R356–R358 (2006).
79. Przeworski, M. The signature of positive selection at randomly chosen loci. *Genetics* **160**, 1179–1189 (2002).
80. Ohyama, H. *et al.* Laser capture microdissection-generated target sample for high-density oligonucleotide array hybridization. *Biotechniques* **29**, 530–536 (2000).
81. Mikulowska-Mennis, A. *et al.* High-quality RNA from cells isolated by laser capture microdissection. *Biotechniques* **33**, 176–179 (2002).
82. Yanai, I. *et al.* Similar gene expression profiles do not imply similar tissue functions. *Trends Genet.* **22**, 132–138 (2006).
83. Gibson, G. & Weir, B. The quantitative genetics of transcription. *Trends Genet.* **21**, 616–623 (2005).
84. Su, A. I. *et al.* Large-scale analysis of the human and mouse transcriptomes. *Proc. Natl Acad. Sci. USA* **99**, 4465–4470 (2002).
85. Su, A. I. *et al.* A gene atlas of the mouse and human protein-encoding transcriptomes. *Proc. Natl Acad. Sci. USA* **101**, 6062–6067 (2004).
86. Altshuler, D. *et al.* A haplotype map of the human genome. *Nature* **437**, 1299–1320 (2005).
87. Lewontin, R. C. & Hubby, J. L. A molecular approach to the study of genic heterozygosity in natural populations. II. Amount of variation and degree of heterozygosity in natural populations of *Drosophila pseudoobscura*. *Genetics* **54**, 595–609 (1966).
88. Darwin, C. *The Origin of Species by Means of Natural Selection; or, the Preservation of Favored Races in the Struggle for Life* (John Murray, London, 1859).
89. Kimura, M. Natural selection as the process of accumulating genetic information in adaptive evolution. *Genet. Res.* **2**, 127–140 (1961).
90. Kimura, M. *The Neutral Theory of Molecular Evolution* (Cambridge Univ. Press, Cambridge, New York, 1983).
91. Ohta, T. Slightly deleterious mutant substitutions in evolution. *Nature* **246**, 96–98 (1973).
92. Ohta, T. Near-neutrality in evolution of genes and gene regulation. *Proc. Natl Acad. Sci. USA* **99**, 16134–16137 (2002).
93. Hudson, R. R., Kreitman, M. & Aguade, M. A test of neutral molecular evolution based on nucleotide data. *Genetics* **116**, 153–9 (1987).
94. McDonald, J. H. & Kreitman, M. Adaptive protein evolution at the *Adh* locus in *Drosophila*. *Nature* **351**, 652–654 (1991).
95. Tajima, F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**, 585–595 (1989).
96. Fu, Y. X. & Li, W. H. Statistical tests of neutrality of mutations. *Genetics* **133**, 693–709 (1993).
97. Fay, J. C. & Wu, C. I. Sequence divergence, functional constraint, and selection in protein evolution. *Annu. Rev. Genomics Hum. Genet.* **4**, 213–235 (2003).
98. Tomita, H. *et al.* Effect of agonal and postmortem factors on gene expression profile: quality control in microarray analyses of postmortem human brain. *Biol. Psychiatry* **55**, 346–352 (2004).
99. 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).
100. Franz, H. *et al.* Systematic analysis of gene expression in human brains before and after death. *Genome Biol.* **6**, R112 (2005).
101. Marvanova, M. *et al.* Microarray analysis of nonhuman primates: validation of experimental models in neurological disorders. *FASEB J.* **17**, 929–931 (2003).
102. 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).
103. Margulies, M. *et al.* Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437**, 376–380 (2005).
104. Whitehead, A. & Crawford, D. L. Variation in tissue-specific gene expression among natural populations. *Genome Biol.* **6**, R13 (2005).
105. Pal, C., Papp, B. & Lercher, M. J. An integrated view of protein evolution. *Nature Rev. Genet.* **7**, 337–348 (2006).

Acknowledgements
The authors thank R.E. Green and S.E. Ptak for many helpful comments on the manuscript and all members of our group for fruitful discussions. The research on primate gene expression in our laboratory is supported by the Max Planck Society and the Bundesministerium für Bildung und Forschung.

Competing interests statement
The authors declare no competing financial interests.

FURTHER INFORMATION
International HapMap project: http://www.hapmap.org
Access to this links box is available online.