MODES OF TRANSCRIPTIONAL REGULATION

Evolutionary dynamics of coding and non-coding transcriptomes

Anamaria Necsulea[1] *and Henrik Kaessmann*[2,3]

Abstract | Gene expression changes may underlie much of phenotypic evolution. The development of high-throughput RNA sequencing protocols has opened the door to unprecedented large-scale and cross-species transcriptome comparisons by allowing accurate and sensitive assessments of transcript sequences and expression levels. Here, we review the initial wave of the new generation of comparative transcriptomic studies in mammals and vertebrate outgroup species in the context of earlier work. Together with various large-scale genomic and epigenomic data, these studies have unveiled commonalities and differences in the dynamics of gene expression evolution for various types of coding and non-coding genes across mammalian lineages, organs, developmental stages, chromosomes and sexes. They have also provided intriguing new clues to the regulatory basis and phenotypic implications of evolutionary gene expression changes.

A primary goal in biology is to understand the molecular basis of phenotypic evolution, most notably that of humans and other mammals. Two major classes of mutations underlie phenotypic innovation: the first comprises mutations that change the coding sequence and consequently the function of the final gene product (that is, the protein or RNA); the second class comprises mutations in regulatory sequences (for example, in promoter regions) that affect transcription, post-transcriptional processing, translation, or transcript and/or protein degradation. Notably, certain gene product sequence alterations that change the function of the protein (for example, mutations in transcription factors[1]) may also have consequences for gene regulation. By modifying developmental programmes, both types of mutations may lead to distinct tissue morphologies, laying the foundation for species- or lineage-specific physiology and behaviour. Moreover, mutations may also directly affect adult organ phenotypes.

The relative importance of the two classes of mutations to adaptation in general, and to morphological and physiological evolution in particular, has been debated since the late 1960s[2–6]. Arguments favouring a dominant role of regulatory change emphasized the potentially deleterious consequences of protein-coding mutations[5,6], which necessarily affect all tissues in which a gene is expressed; such arguments also highlight that coding sequence differences between closely related yet phenotypically disparate species (for example, humans and chimpanzees) are rare[2]. Counterarguments focused on individual cases in which protein structural mutations contributed to organismal adaptation[4].

However, only the rise of the genomic era in the past decade has begun to provide a more solid foundation for a systematic exploration of the molecular basis of phenotypic evolution. Comparisons of a rapidly increasing number of mammalian genome sequences have revealed numerous selectively driven amino acid substitutions[7] and even potentially adaptive promoter changes[8]. Moreover, the development of large-scale technologies has facilitated genome-wide assessments of expression patterns, as well as of *cis*- and *trans*-acting regulatory mechanisms. Pioneering microarray-based transcriptome comparisons between closely related mammals identified potentially adaptive expression changes and provided initial evidence for principles governing the evolution of gene expression[9,10]. However, the application of microarrays for interspecies comparisons is limited (BOX 1). Consequently, the recent advent of high-throughput sequencing approaches[11], which enabled detailed qualitative and quantitative comparisons of transcriptomes and gene regulatory mechanisms between divergent species, represented a breakthrough in the field of molecular evolution.

Here, we discuss the exciting insights that these novel approaches have already provided into the dynamics and phenotypic implications of transcriptome evolution. Our Review differs from previous accounts[12] in that we

[1]*Laboratory of Developmental Genomics, School of Life Sciences, École Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland.*
[2]*Center for Integrative Genomics, University of Lausanne, 1015 Lausanne, Switzerland.*
[3]*Swiss Institute of Bioinformatics, 1015 Lausanne, Switzerland.*
e-mails: anamaria.necsulea@epfl.ch; henrik.kaessmann@unil.ch
doi:10.1038/nrg3802
Published online 9 October 2014

Cis- and *trans*-acting regulatory mechanisms
Mechanisms that regulate gene expression. *Cis*-acting regulatory elements (such as promoters or enhancers) are DNA sequences located on the same chromosome and generally in the vicinity of the target gene, and they recruit molecules that can modulate the expression of the target gene. *Trans*-acting factors (such as transcription factors or microRNAs) control gene expression by binding to *cis*-acting elements; their genomic localization is independent of that of the regulated genes.

provide a cross-mammalian (vertebrate) view, traverse both the coding and the non-coding transcriptome, and cover recent studies that unveiled previously unknown aspects of gene expression evolution. We mainly discuss long-term evolutionary changes of the transcriptome, with brief references to intraspecies variations (which have been reviewed elsewhere[13,14]).

We set the stage by presenting known components of the transcriptome and proceed by examining their rates of expression changes in different mammalian lineages and organs. We also highlight patterns of sex-related expression remodelling and summarize our current knowledge concerning the phenotypic relevance of transcriptome changes. We conclude our account by discussing recent insights into the regulatory basis of transcriptome evolution and identifying promising future research avenues.

Transcriptome complexity

Until recently, most transcriptomic studies focused on protein-coding genes. However, the recent development of dedicated technologies (BOX 1) has enabled more comprehensive transcriptome investigations. High-throughput RNA sequencing (RNA-seq) has revealed that most mammalian protein-coding genes have multiple splicing variants, many of which are expressed in a tissue-specific manner[15]. Numerous isoforms are also generated through differential usage of promoters, transcription start sites[16] and polyadenylation sites[17]. The evolution of the protein-coding fraction of the transcriptome can thus occur at multiple levels.

In addition to protein-coding mRNAs, mammalian genomes encode various non-coding RNAs. Some of these transcripts participate in core functions such as splicing (small nuclear RNAs), mRNA translation

Box 1 | Evolutionary transcriptomics studies powered by next-generation sequencing technologies

Large-scale transcriptomic studies are rapidly gaining power owing to the development of new technologies. Microarrays brought the first revolutionary step in transcriptomics by allowing large-scale quantitative assessments of gene expression[132] at a low cost. They were successfully applied to compare protein-coding gene expression levels between tissues or cell types, developmental stages or disease associations within a single species[133]. However, as they rely on hybridization to a priori designed probes, microarrays prohibit the discovery of new genes or new alternative transcripts and are thus mainly restricted to the study of known genes (although more recently tiling arrays have enabled genome-wide investigations of transcribed regions[134]). Moreover, their application to evolutionary studies is limited, as cross-species hybridization to the same set of probes is confounded by sequence divergence, whereas the use of species-specific probes may not yield comparable results[135]. Thus, microarray-based studies have been confined to closely related species such as primates[9,10,136] or mouse species[9], although a comparison among vertebrates was attempted[37].

In the past few years, the use of microarrays for evolutionary transcriptomics has been largely abandoned in favour of more powerful technologies based on high-throughput sequencing. RNA sequencing (RNA-seq) has numerous advantages over microarrays even for intraspecies transcriptomic studies, as it enables a wider dynamic range for transcript detection[101], provides a better quantification of expression levels and more comparability with proteomics studies[137], and permits *de novo* identification of genes and isoforms[138], as well as gene expression quantification even in the absence of genomic data[139]. RNA-seq has proved to be an excellent tool for evolutionary transcriptomics by enabling comparisons across distantly related species for numerous aspects of the transcriptome. So far, this technology has been used for large-scale evolutionary assessments of protein-coding gene expression levels[39,139], alternative splicing patterns[43,44], microRNA (miRNA)[46,140] and long non-coding RNA (lncRNA) repertoires and expression patterns[50–52], and RNA editing levels[141,142]. Other techniques based on high-throughput sequencing have been used to analyse patterns of mRNA polyadenylation and promoter or transcription start site usage across species[16,45].

Further technological innovations that can be exploited for evolutionary transcriptomics studies are expected in the near future. In particular, single-cell RNA-seq technologies[128] may help to assess the impact of cell population heterogeneity on interspecies transcriptome comparisons.

The table below summarizes the usage of different molecular techniques for evolutionary transcriptomics studies and provides a non-exhaustive list of representative publications.

Technology	Transcriptome aspects detected	Species studied
Hybridization-based approaches		
Microarrays	Protein-coding gene expression levels	Great apes[9,10], primates[136], mouse species[9] and vertebrates[37]
Tiling microarrays	Intergenic transcription levels	Humans and chimpanzees[134]
Exon junction microarrays	Alternative splicing	Humans and chimpanzees[42], and primates[143]
High-throughput sequencing		
RNA-seq	Protein-coding gene expression levels	Amniotes[39] and primates[139]
	miRNA expression	Amniotes[46] and primates[121,140]
	Alternative splicing	Primates[34], mammals[43] and tetrapods[44]
	lncRNA expression	Rodents[50], mammals[52] and tetrapods[51]
	mRNA editing	Primates[141]
Poly(A) sequencing	Alternative polyadenylation	Mammals[45]
Cap analysis gene expression (CAGE)	Promoter and transcription start site usage	Humans and mice[16]
Digital expression tag profiling (TAG-seq; also known as DGE-TAG)	Protein-coding and non-coding RNA expression levels	Primates[144]

(ribosomal RNAs and tRNAs) or post-transcriptional rRNA editing (small nucleolar RNAs). MicroRNAs (miRNAs), which attenuate gene expression by promoting mRNA degradation or translational repression[18], are also essential molecules. Another class of short non-coding RNAs, PIWI-interacting RNAs (piRNAs), silence transposable elements in the germ line and may regulate gene expression in the soma[19]. These different transcript categories can be assayed using dedicated techniques, which select specific RNA size ranges or features (for example, poly(A) tails). Here, we mainly address the evolution of polyadenylated transcripts, including protein-coding mRNAs, most long non-coding RNAs (lncRNAs) and miRNA precursors.

Recently, lncRNAs have attracted considerable attention. Although several lncRNAs have been known for decades[20], it was only recently established that mammalian genomes encode thousands of lncRNAs[21–23]. Most lncRNAs have not yet been functionally characterized, but individual lncRNAs have been implicated in gene expression regulation. These include *cis*-acting silencers — for example, the well-known lncRNA X inactive specific transcript (*XIST*), which underlies X chromosome inactivation in females[24], and the imprinted lncRNA antisense of IGF2R non-protein coding RNA (*AIRN*)[25] — *cis*-acting enhancers[26], *trans*-acting regulators that recruit chromatin-modifying complexes to specific genomic locations[27] and competing endogenous RNAs that act as miRNA 'sponges' (REF. 28). Regulatory non-coding RNAs (including lncRNAs, miRNAs and piRNAs) are important for evolutionary transcriptomics studies, as they are likely to be part of the mechanisms underlying expression evolution.

The complexity of mammalian transcriptomes is further illustrated by the frequent overlap between different transcript categories or between genes and other genomic elements. For example, lncRNAs can act as miRNA precursors[29], and enhancer activity is often accompanied by transcription of non-coding RNA[30,31]. Moreover, during evolution, the boundaries between different gene categories can be blurred, as illustrated by the emergence of the *XIST* lncRNA through pseudogenization of a protein-coding gene[32]. These aspects of the transcriptome have only recently begun to be investigated at a large scale, and future studies are likely to reveal further complexity.

Evolutionary rates of various transcriptome parts

Protein-coding transcripts. Although many evolutionary studies focus on identifying lineage-specific changes, a striking observation of comparative transcriptomics is the overall high level of expression conservation, at least for protein-coding genes. Protein-coding gene expression was shown to display low variation across primates, probably as a result of negative selection (that is, purifying selection)[10,33,34]. For example, absolute expression differences between human and chimpanzee livers were on average only 20% higher than differences observed between pairs of individuals of the same species for this tissue[34]. Early microarray-based analyses between distant species (humans and mice) suggested rapid (that is, essentially neutral) evolution of gene expression[35], but this was later disproved with thorough statistical corrections for variability between the probes used[36]. A cross-vertebrate analysis also indicated strong expression conservation, even in the absence of constrained exonic sequences[37]. A comparison across distantly related amniote species and multiple major organs confirmed that protein-coding gene expression divergence does not accumulate linearly with time (as would be expected under neutral evolution[38]). Rather, expression patterns are similar even between divergent species such as humans and chickens (with Spearman's rank correlation coefficients above 0.7 for somatic organs), suggesting that conservation of core organ functions limits transcriptome divergence[39]. Thus, for protein-coding gene expression levels, the main source of variability is the tissue in which they are measured: different samples cluster together on the basis of the organ from which they are derived, irrespective of the species of origin[39] (FIG. 1a).

Alternative isoforms. Divergence rates are much higher for other aspects of the transcriptome, particularly for alternative splicing patterns, than for protein-coding gene expression levels. An initial comparison of exon skipping frequencies between humans and mice reported high evolutionary conservation[40], but this observation was later contradicted by more comprehensive human–mouse and human–chimpanzee comparisons, which revealed frequent species-specific alternative splicing events[41,42]. The pervasiveness of species-specific alternative splicing was recently confirmed in cross-mammal and tetrapod studies based on RNA-seq data for multiple organs[43,44]. This rapid evolution is illustrated by a species-dominated pattern of clustering: exon skipping frequencies are more similar between different organs of the same species than between different species for a given organ[43,44] (FIG. 1b), which is in contrast to the grouping observed for protein-coding gene expression levels (FIG. 1a). The high species specificity of alternative splicing was convincingly shown to be driven by changes in *cis*-acting regulatory elements, on the basis of the finding that a mouse model carrying a copy of human chromosome 21 shows human-specific exon skipping frequencies on the inserted chromosome (mouse-specific patterns would be expected if *trans*-acting regulatory factors were predominant[44]).

Less is known about other mechanisms that generate isoform diversity. A cross-mammal comparative study of polyadenylation site usage indicated that the absolute usage frequencies of polyadenylation sites are better correlated between samples derived from the same organ from different species, rather than between samples from different organs within a single species[45], as observed for protein-coding gene expression levels (FIG. 1a). This indicates that absolute usage frequencies of these sites are generally well conserved across species. However, it is not yet clear whether the pattern of alternative polyadenylation site usage (for genes with multiple polyadenylation sites) shows the same extent of conservation, or whether it displays the rapid evolution characteristic of alternative splicing.

Negative selection
A type of natural selection that eliminates deleterious mutations.

Amniote
A subclass of vertebrate species characterized by the presence of an amnion, which is a protective envelope that encloses the embryo. They include reptiles, birds and mammals but not amphibians.

Spearman's rank correlation coefficients
In transcriptome analyses, non-parametric statistics (in which the data are not required to follow specific statistical distributions) are often preferred. For example, expression levels may be compared across samples using Spearman's rank correlation coefficient, a statistical measure that estimates the similarity between the ranks of the genes, rather than between their absolute expression values.

Exon skipping
A form of alternative splicing in which internal exons are alternatively included or excluded (skipped) in the mature mRNA.

Non-coding transcripts. The non-coding complement of the transcriptome includes both highly conserved and rapidly evolving categories of genes. For miRNAs, the ~100 genes found to be shared across mammals are slow-evolving in terms of DNA sequence[46] and expression (FIG. 1c). Expression levels of these conserved miRNAs are highly correlated across distantly related species and different organs, perhaps even more than what is observed for protein-coding genes (FIG. 1a,c). However, part of this conservation may be due to biases associated with commonly used approaches for small-RNA sequencing, which strongly favour the representation of some miRNAs over others[47,48]. miRNA repertoires also include numerous lineage-specific genes, which are generally expressed at lower levels[46,49].

lncRNA repertoires undergo rapid turnover[50–52], and lncRNA sequences and expression levels tend to evolve more rapidly[50,51] than those of protein-coding genes, although hundreds of lncRNAs that are highly conserved at either or both levels have been identified[51,53]. Even for lncRNAs that are shared across multiple species, the extent of expression level conservation is much lower than that observed for protein-coding genes[51]. However, organ specificity of lncRNAs is more conserved among species than expected by chance[51,52], although the level of conservation is still less than that of protein-coding genes[51]. Part of the observed interspecies expression differences may stem from the typically low abundance of lncRNA transcripts, which adds 'noise' to expression estimates[51]. The weak lncRNA sequence conservation may also inflate estimated expression divergence for distant species, for which homologous exonic sequences are difficult to identify[51]. These technical biases lead to overestimation of lncRNA expression divergence, although rapid lncRNA evolution is undoubtedly real[51]. Overall, comparisons of lncRNA expression levels reveal no clear clustering pattern (FIG. 1d), which reflects the presence of both fast-evolving and conserved organ-specific genes. This pattern

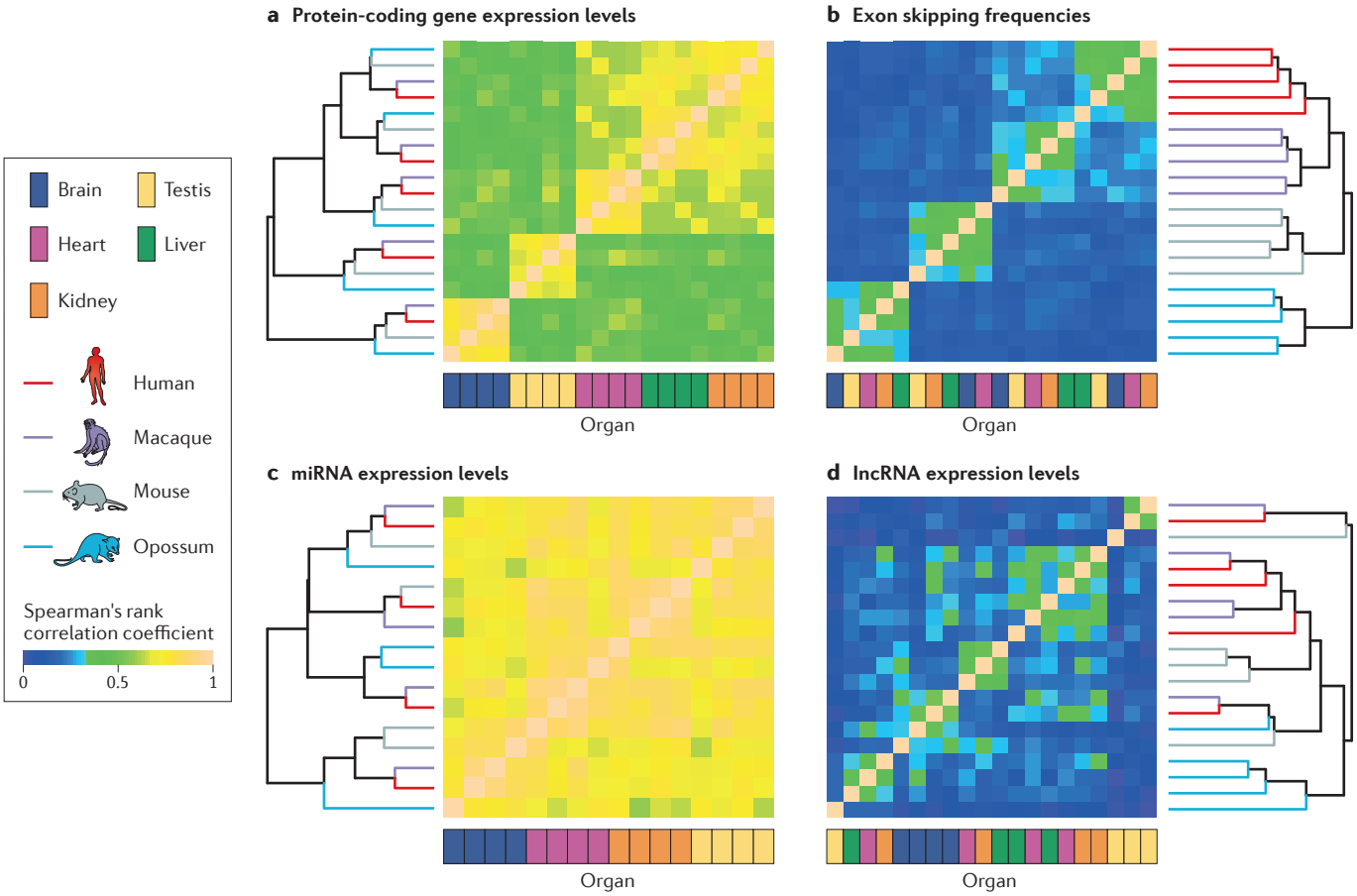


Figure 1 | **Global patterns of evolution for different aspects of the transcriptome. a** | Hierarchical clustering of protein-coding gene expression levels, estimated as the number of reads per kilobase of exon per million mapped reads (RPKM), is shown. The heatmap represents Spearman's rank correlation coefficients between pairs of samples. The sample clustering is represented as a tree with branch colours depicting different species. **b** | Hierarchical clustering shows that exon skipping frequencies are more similar between different organs of the same species than between different species for a given organ[43,44]. Exon skipping frequencies were estimated on the basis of spliced reads counts as described in REF. 15. Processed RNA-seq data in parts **a** and **b** were taken from REF. 39. **c** | Hierarchical clustering of microRNA (miRNA) expression levels, estimated as the number of uniquely mapped reads per miRNA gene, shows that miRNA expression levels are highly correlated across organs and species. RNA-seq data were taken from REF. 46. **d** | Hierarchical clustering of long non-coding RNA (lncRNA) expression levels (RPKM values) is shown. Processed RNA-seq data were taken from REF. 51.

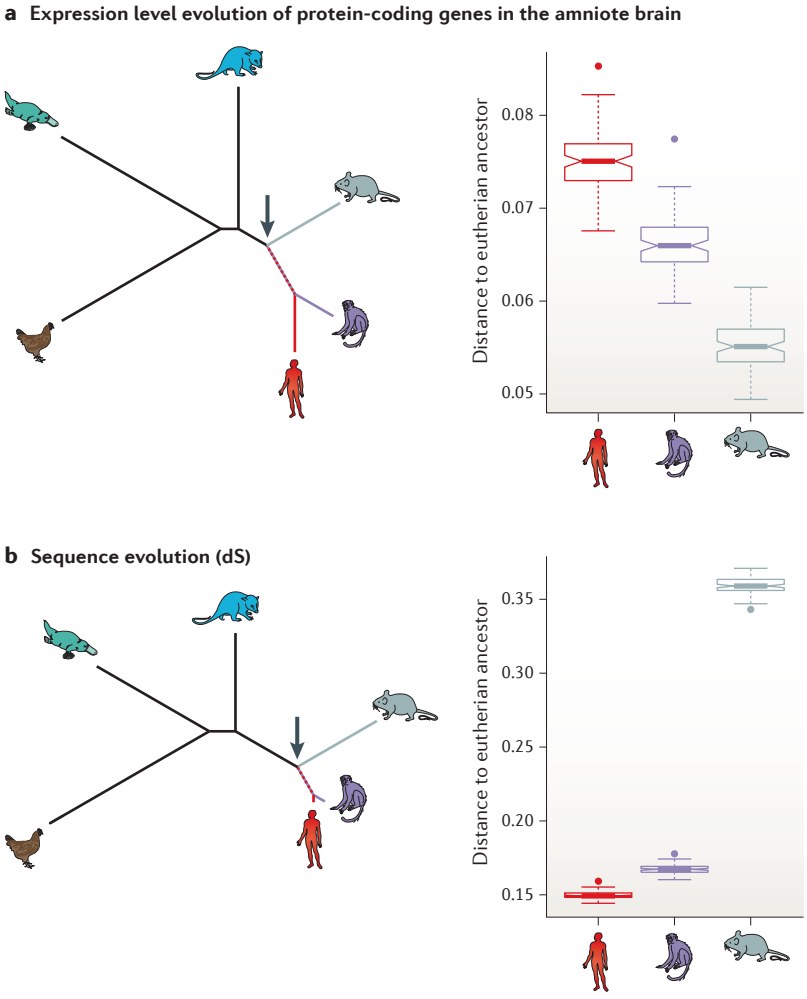


Figure 2 | **Rates of genome and transcriptome evolution in different mammalian lineages. a** | Neighbour-joining tree of protein-coding gene expression levels in the amniote brain is shown (left panel). The tree was constructed on the basis of pairwise distances between samples, estimated as one minus the Spearman's rank correlation coefficient. Colours highlight branches leading from the human, macaque and mouse ancestor (indicated by the arrow) to each of the three species. Expression data were taken from REF. 39. Distance between each of the three placental mammal species and their common ancestor in the expression tree is shown in the boxplot (right panel). The boxplot represents the distribution of the distances to the ancestor for humans, macaques and mice, observed in 100 bootstrap replicates, in which the same numbers of genes were resampled with replacement. The medians are shown in the middle of each box by the bold lines. The whiskers of the boxplots extend to the most extreme data point, which is no more than 1.5 times the interquartile range away from the box. The notches indicate 95% confidence intervals for the median. **b** | Rates of synonymous substitutions in protein-coding sequences (dS) are depicted on an amniote species tree (left panel). Branch lengths are proportional to lineage-specific dS rates. dS values were estimated with PAML[145] on concatenated multiple alignments of 2,000 randomly selected homologous protein-coding sequence families extracted from the Ensembl database (release 75)[146]. Distance between each of the three placental mammal species and their common ancestor (indicated by the black arrow) in the dS tree is shown in the boxplot (right panel). The boxplot represents the distribution of the dS rates observed with 100 resampling replicates, in which dS levels are computed on different sets of 2,000 protein-coding gene families.

is likely to be influenced by the noisy lncRNA expression estimates, and indeed with greater RNA-seq depth the organ-dependent clustering is predominant[52].

Little is known about the patterns of expression evolution of other categories of non-coding RNAs, although one study indirectly assessed patterns of tRNA expression by examining frequencies of RNA polymerase III binding[54], which revealed surprisingly high evolutionary divergence for individual tRNA genes yet high conservation for the usage of families of tRNAs with the same anticodon[54].

Rates of evolution in different lineages

Comparative transcriptomics analyses have revealed that mammalian lineages have undergone different rates of expression evolution[9,10,39]. For example, gene expression levels evolve faster in primates than in rodents[39] (FIG. 2a). Primates are also characterized by a faster accumulation of alternatively spliced isoforms than other mammalian lineages[43]. This primate-specific acceleration of transcriptome evolution cannot be explained by mutation rate differences, given that rodents have much higher mutation rates than primates owing to their short generation time[55]. Indeed, as a consequence of this elevated mutation rate, synonymous substitution rates are considerably higher in the mouse lineage than in primates (FIG. 2b). Thus, in the absence of natural selection, elevated mutation rates would increase the divergence of regulatory sequences and thus accelerate expression evolution in rodents.

For the brain, it is tempting to explain the primate-specific acceleration of expression evolution by invoking adaptive changes underlying the evolution of the higher complexity of this organ in primates. However, the acceleration was also observed for all other studied organs[39], in which massive primate-specific adaptations are less expected. Moreover, other lineages (for example, monotremes) also show a significant increase in the rate of expression evolution compared with rodents[39]. On the basis of these observations, population genetics reasoning suggests that the adaptive scenario is unlikely and that another evolutionary parameter may provide a better explanation: the effective population size (N_e). The long-term N_e of rodents[56] has been estimated to be at least ten times higher than that of humans (10,000–20,000 individuals[57,58]) and chimpanzees (~35,000 individuals[59]). Similarly, the platypus has lower N_e than rodents[60]. The lower N_e in primates and monotremes, which implies weaker efficiency of natural selection, leads to a higher fixation rate for mildly deleterious mutations that affect gene expression levels. Thus, faster expression evolution in these lineages is more likely to be explained by less efficient selection rather than by more frequent adaptive changes, in agreement with an analysis of regulatory sequence evolution, which revealed an accelerated degradation of hominid regulatory elements that is probably due to accumulation of slightly deleterious mutations[56]. Evolutionary rate differences among lineages are thus consistent with a nearly neutral model of expression evolution, in which most mutations affecting gene expression are (slightly) deleterious[61,62].

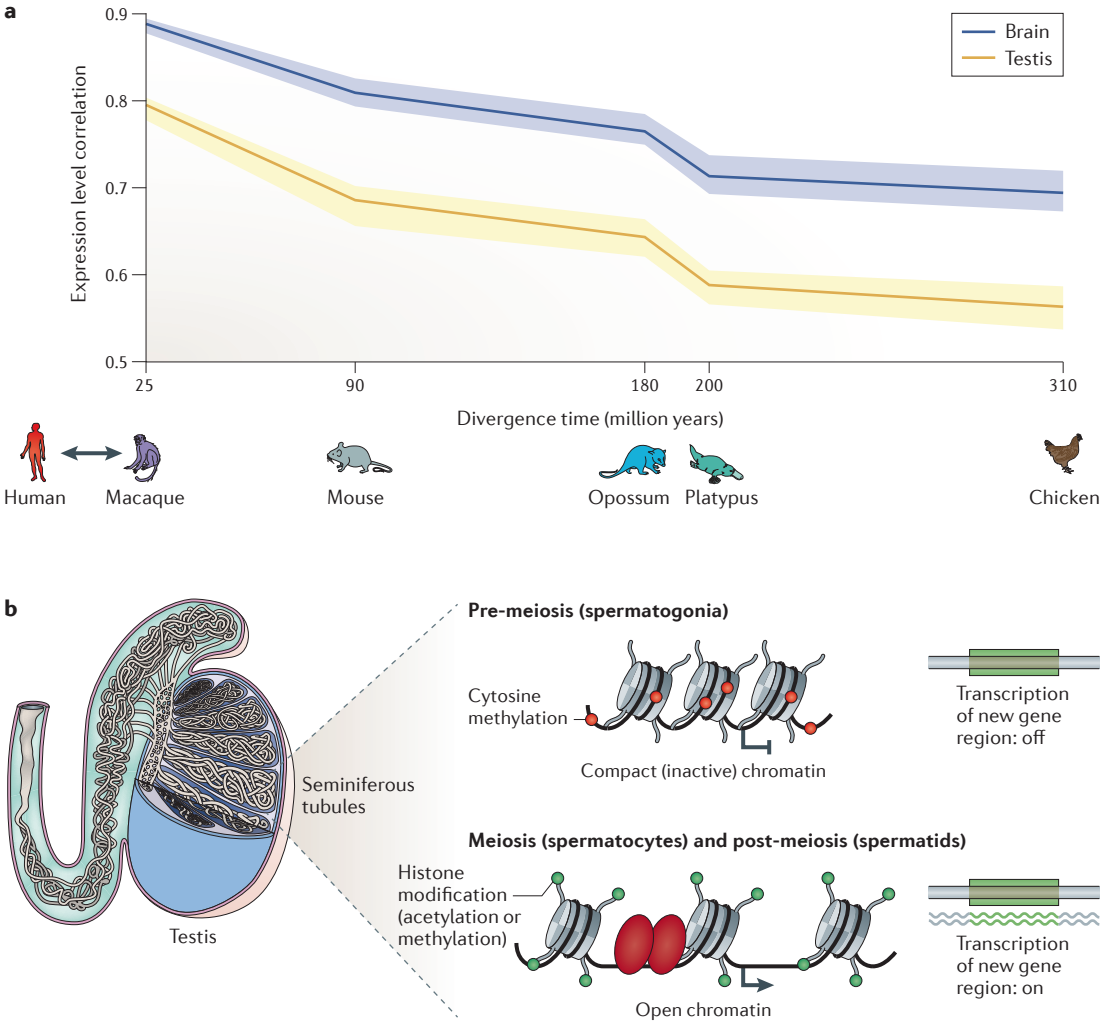


Figure 3 | **Different rates of evolution in major mammalian organs and 'promiscuous' transcription in the testis.** **a** | Expression levels evolve more slowly in neural tissues than in the testes. Pairwise Spearman's rank correlation coefficients between humans and other species are shown as a function of the divergence time between the two species for brain and testis protein-coding gene expression levels. Expression data were taken from REF. 39. **b** | Transcription in meiotic spermatocytes and post-meiotic spermatids (germ cells), which dominate total testis expression signals, is widespread owing to an overall permissive chromatin conformation that results from extensive repackaging of DNA during spermatogenesis[66]. Chromatin in pre-meiotic spermatagonia is in a compact (inactive) state but changes to an overall open (relaxed) state in meiotic spermatocytes and post-meiotic (round) spermatids as a consequence of extensive remodelling events during these stages of spermatogenesis[66], which take place in the seminiferous tubules. The transcriptionally permissive chromatin conformation is achieved through widespread demethylation of CpG dinucleotide-enriched promoter sequences and by histone modifications (acetylation and methylation), which facilitates access of the transcriptional machinery (red ovals). The specific chromatin environment during and after meiosis promotes the initial transcription of newly emerged genes during evolution[66].

Rates of expression changes in different organs

Rates of transcriptome evolution also vary among organs, which indicates global differences in selective pressures. Expression levels evolve more slowly in neural tissues than in other organs in vertebrates[10,37,39,51]; this pattern holds true for both protein-coding[10,37,39] and lncRNA genes[51], and thus reflects a transcriptome-wide trend (FIG. 3a). The slow expression evolution in nervous tissues is paralleled by low rates of protein sequence divergence of brain-specific genes[10,63]. These observations are notable when considering the pronounced remodelling of the size, structure and cellular composition of the brain during evolution[39], as well as the difficulty in sampling and comparing corresponding brain regions across species[39], which may have exaggerated actual biological expression differences. Overall, the brain may thus have particularly sensitive and 'fine-tuned' regulatory networks that are highly constrained during evolution. Nevertheless, certain network components may sometimes be rapidly reshaped during evolution, for example, through (developmental) expression

Effective population size (N_e). A measure of the genetic diversity of a population. It is typically defined as the number of individuals in an idealized population (that is, one with random mating, equal sex distribution, etc.) that would show the same amount of genetic diversity as the population under consideration.

changes of a few key regulators[64] (such as miRNAs). These events may provide an explanation for major phenotypic brain innovations, such as the sudden and pronounced increase of its size and complexity in the human lineage[64].

The testis stands out at the other end of the spectrum with a particularly high rate of transcriptome evolution[10,39,51], presumably for two major reasons. First, the testis underwent the largest number of shifts in gene expression levels during mammalian evolution[10,39], which is consistent with the rapid protein-coding sequence divergence of testis genes[10,65] and probably reflects the intense sex-related selective pressures (for example, those associated with sperm competition) acting on this organ[65]. Second, evolutionary constraints on transcription may be globally reduced in the testis. Indeed, it was recently shown that transcription in certain abundant germ cells (meiotic spermatocytes and round spermatids), which dominate total testis expression signals, is widespread and partly 'promiscuous' owing to an overall permissive chromatin conformation that results from extensive repackaging of DNA during spermatogenesis[66] (FIG. 3b). Therefore, relaxation of purifying selection, because of relaxed transcriptional regulation during and after meiosis, probably also contributes to the rapid evolution of testis transcriptomes[66]. Thus, many testis-specific transcripts, in particular evolutionarily recent lncRNAs, may represent transcriptional noise. The peculiar chromatin environment during spermatogenesis is likely to have catalysed the emergence of new coding[67,68] and non-coding[51,67] genes during evolution by facilitating their initial transcription (the 'out of the testis' scenario[51,67]).

We note that most previous evolutionary studies sought to assess transcriptome patterns in samples representing whole organs, with a notable exception for the brain[69], which revealed slower evolution of neurons than endothelial cells in humans and chimpanzees. It will be rewarding to compare transcriptomes for individual cell types in the future, particularly for complex and heterogeneous tissues. This work will help to resolve remaining questions about the evolutionary dynamics of transcriptomes, such as whether previously measured brain expression shifts in the primate lineage reflect actual changes in cellular transcript abundance or whether such shifts constitute innovations in terms of cellular composition of tissues.

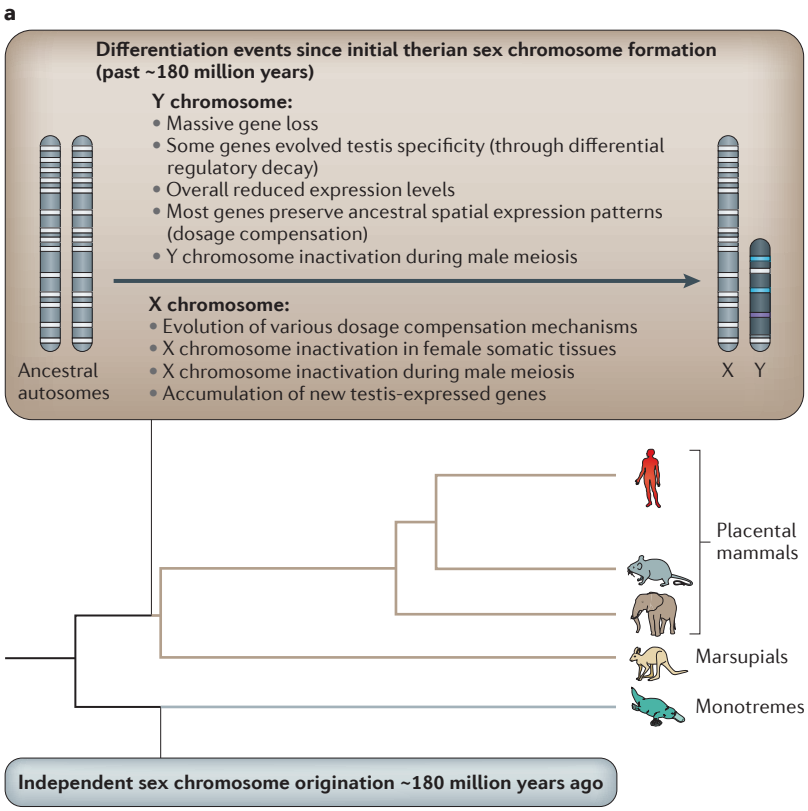
Sex-related transcriptome evolution

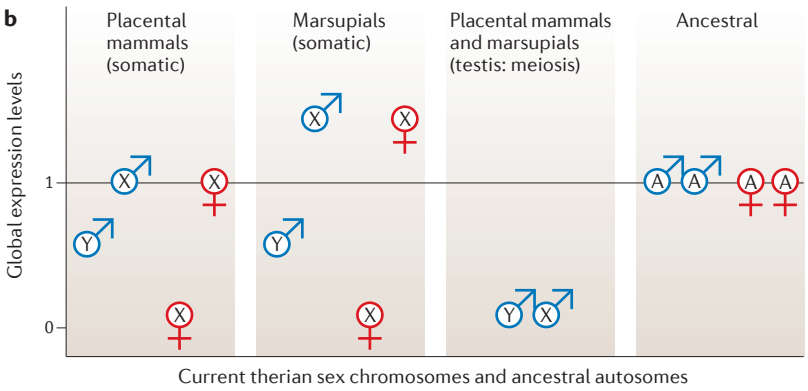
Autosomal sex biases. Sex-biased gene expression in mammals is not restricted to sex-specific organs, such as the testis (discussed above) or the ovary, but is also widespread in somatic tissues, with thousands of genes showing sexually dimorphic patterns[70]. However, expression level differences between the sexes are commonly weak, at least on autosomes, and therefore usually require large sample sizes for detection[70]. Nevertheless, evolutionary studies uncovered sex-specific expression signatures on autosomes, some of which are shared across species[39,71].

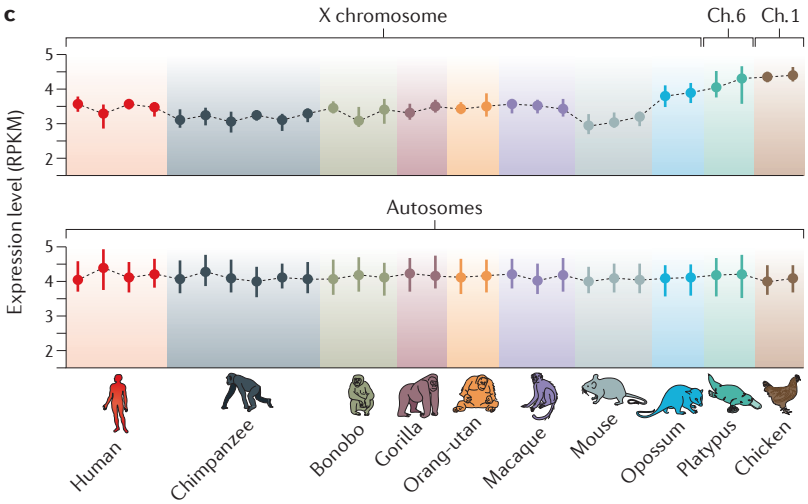
Therians
A subclass of mammals that comprises placental mammals (eutherians) and marsupials (metatherians); that is, mammals that give birth to live young without using a shelled egg. The remaining major mammalian lineage is that of the egg-laying monotremes (prototherians).

Figure 4 | **Origins and functional evolution of mammalian sex chromosomes.** **a** | The independent origins of sex chromosomes in the ancestor of therians (that is, placental mammals and marsupials) and monotremes are shown, along with an overview of major gene content and expression remodelling events during therian sex chromosome evolution. **b** | Data from recent studies[73,77] revealed the overall patterns of the evolution of global expression levels of the X and Y chromosomes in males and females in somatic tissues and the testis. The plot displays current expression levels of X- and Y-linked genes in extant therian mammals and those of ancestral autosomal (proto-sex chromosomal) precursors from which current X- and Y-linked genes are derived. Ancestral expression levels of autosomal precursors are inferred from expression levels of orthologues of X- and Y-linked genes in species with different sex chromosome systems[73,77] (for example, platypus and chickens), in which these genes are still autosomal. Inferred proto-sex chromosomal expression levels are set to one in this plot; all other expression levels are plotted relative to this ancestral expression level (zero indicates no expression). **c** | A cross-species comparison of global chromosomal expression outputs in the brain is shown. The upper panel shows the overall reduced expression output of genes on the single active X chromosome in therians (particularly in placental mammals) relative to that of corresponding autosomal genes (two copies) in outgroup species (that is, chromosome 6 (Ch. 6) genes in platypus and chromosome 1 genes in chickens), which approximately reflect the ancestral expression output of the autosomes from which current therian sex chromosomes are derived[73,77]. Thus, this plot illustrates the substantially (on average twofold) reduced expression output of the current X chromosome in placental mammals compared with ancestral autosomes, which suggests an overall lack of transcriptional upregulation to compensate for the massive loss of Y-linked genes. Individual genes were probably nevertheless transcriptionally upregulated, and other mechanisms also compensated for this dosage reduction. The displayed expression levels for each individual are based on third quartiles of the distributions of genes (in reads per kilobase per million mapped reads (RPKM)) on the different chromosomes; error bars represent the range containing 90% of the third quartiles of individual resampling sets. A control analysis of autosomal genes in all species (lower panel) shows that global autosomal expression outputs have remained essentially unaltered during evolution. Part **c** adapted from REF. 77.

Sex chromosome origins and functional remodelling. Sex chromosomes, which are genomic hot spots of sex-related expression evolution, arose twice independently in mammals[72,73]. The original sex chromosomes of placental mammals and marsupials (collectively referred to as therians) originated in the therian ancestor ~180 million years ago, in parallel with monotreme sex chromosomes[73] (FIG. 4a). The differentiation of these sex chromosomes, which are derived from ordinary pairs of ancestral autosomes, entailed substantial chromosome-wide remodelling of gene contents and expression patterns due to sex-related selective forces (FIG. 4a) (see below).







Functional evolution of Y chromosome genes. The specialization of the Y chromosome on sex determination and male functions, together with recombination arrests that weakened selection efficiency, led to a drastic reduction of Y-linked genes during mammalian evolution[73] (FIG. 4a). Some remaining Y-linked genes evolved testis-specific expression patterns[73], which is consistent with the male-limited transmission of this chromosome. Interestingly, these genes, which were originally widely expressed across tissues on the ancestral proto-sex chromosomes, evolved testis specificity through differential regulatory decay; that is, they underwent substantially stronger expression reduction in somatic tissues than in the testis during evolution[73]. However, most genes on the Y chromosome seem to have general regulatory functions (that is, they may regulate transcription, splicing, translation and stability of encoded proteins of other genes) and maintained ancestral ubiquitous expression patterns, although their expression levels decreased[73] (FIG. 4b). Together with other observations (for example, that regulatory genes are commonly dosage sensitive), it was concluded that most genes on the Y chromosome were initially preserved because of dosage constraints in order to maintain, in conjunction with their counterparts on the X chromosome, the ancestral gene copy number, thus also ensuring proper functioning of male somatic tissues[73,74].

Evolution of X chromosome gene dosage compensation. However, how was the massive loss of all other genes from the Y chromosome compensated? The prevailing theory posits that a twofold upregulation of the single remaining gene copies on the X chromosome generally restored ancestral expression outputs[75,76]. The overabundance of X-linked transcripts in females, which results from the combined activity of the two upregulated X chromosomes, was then compensated by the inactivation of one of the X chromosomes. Yet, surprisingly, it was recently shown that the overall expression output of X-linked genes is only half of that inferred for proto-sex chromosomes, at least in placental mammals[77,78] (marsupials do show signs of global upregulation of the active X chromosome in several tissues[77,78]) (FIG. 4b, c). Nevertheless, individual X-linked genes may have become actively upregulated in placental mammals through specific mechanisms[77,79]. Moreover, the dosage reduction of some X-linked genes was compensated by a corresponding twofold downregulation of autosomal partner genes, which maintained proper cellular balances of interacting gene products[77]. Other mechanisms may have also contributed to X chromosome dosage compensation[77], and various genes may have been insensitive to dosage alterations altogether[77]. Nonetheless, the reason underlying the evolution of chromosome-wide X chromosome inactivation in females (FIG. 4a, b) remains overall enigmatic in view of these findings.

X-linked gene expansions and sex chromosome inactivation in males. Early X chromosome evolution was accompanied by increased rates of functional adaptation, as indicated by accelerated expression evolution

Dosage sensitive
Pertaining to genes for which the associated phenotypes are directly affected by modifications of their transcriptional output (such as changes in the number of active copies in the genome). For example, genes that encode subunits of a protein complex are dosage sensitive, as individual gene duplication leads to an imbalance among the parts of the complex.

X chromosome dosage compensation
A process that compensates for the loss of most Y-chromosome genes after sex chromosome differentiation from ancestral autosomes (that is, the reduction from two copies on sex chromosome precursors to one gene copy on the X chromosome in males).

Positive selection
A type of natural selection that favours the spread to fixation of an allele that increases the fitness of individuals.

Hemizygous
A condition in which there is only one copy of a gene in an otherwise diploid cell or organism (for example, all genes on the X chromosomes of male mammals).

Meiotic sex chromosome inactivation
(MSCI). A process leading to the transcriptional silencing of the X and Y chromosomes during the meiotic phase of spermatogenesis.

Genetic drift
An evolutionary process through which the frequency of an allele in a population changes through time simply by a random sampling effect; that is, some alleles can increase or decrease in frequency by chance over time, even in the absence of an effect on the fitness of the individuals. The magnitude of this phenomenon is stronger for lower effective population sizes.

Ornstein–Uhlenbeck processes
Mathematical models that describe random variations of a quantitative variable towards an optimum value. This model is widely used to describe evolutionary processes, as it can incorporate both genetic drift and natural selection.

during this time that may reflect the evolution of dosage compensation (see above) and novel functions of pre-existing genes[39]. However, the evolution of the newly formed X chromosome was also accompanied by an increased fixation rate of duplicate gene copies[46,77,80] (FIG. 4a). In particular, both coding[77,80] and non-coding genes[46] with testis expression massively expanded on the X chromosome through gene duplication upon sex chromosome differentiation, and this was probably facilitated by the specific selective environment on the X chromosome[81], in which male-beneficial mutations are always visible to positive selection because of the single-copy (that is, hemizygous) status of the X chromosome in this sex. However, the amplification of these genes was probably also necessary to evade meiotic sex chromosome inactivation (MSCI)[82] in males and its post-meiotic consequences[46,83]. Other X-linked genes with functions in spermatocytes and spermatids escaped MSCI through an alternative mechanism that involves the export of duplicate copies, which have specific transcriptional activity in these cells, to autosomes (the 'out of the X' pattern[67]).

Phenotypic relevance of transcriptome changes

Identifying phenotypically relevant changes. Several experimental studies demonstrated that gene expression evolution may indeed have marked phenotypic effects. For example, changes in the developmental expression pattern of the bone morphogenetic protein 4 (*BMP4*) gene were found to induce the particular broad beak morphology of Darwin's finches[84], whereas an increase in the expression level of this gene in the developing genital tubercle is responsible for the evolutionary reduction of the phallus in galliform birds[85]. Similarly, changes in the expression pattern of the agouti gene cause coat colour differences between different populations of *Peromyscus* deer mice[86], and changes in the expression of the KIT ligand (*KITLG*) gene were shown to underlie blond hair colour in Europeans[87]. However, experimental evidence for causal relationships between expression changes and phenotypic alterations remains rare, probably owing to the complexity of the required experiments, but perhaps also because of the difficulty in identifying the best candidate genes. This emphasizes the need for evolutionary inference of potential phenotypically relevant expression changes that can be functionally characterized. Thus, numerous evolutionary studies have attempted to identify adaptive changes in gene expression, which are likely to be associated with phenotypic innovations.

Several mathematical frameworks have been developed to determine the selective pressures that act on gene expression[33,34,39,88–90]. Similar to classical approaches for investigating the selective pressures acting on DNA sequences, these methods generally require knowledge of intraspecific and interspecific gene expression variation. Specifically, high ratios of intraspecies diversity over interspecies divergence indicate that most mutations affecting gene expression are deleterious, and that gene expression evolves under purifying selection. Thus, the gene can also be said to evolve under stabilizing selection, which preserves its optimal expression level and reduces the variation around it. Conversely, an excess of interspecies divergence compared with intraspecies diversity is suggestive of positive selection, which directs the expression level towards a new (lineage-specific) optimum value. These principles have been implemented in several statistical methods, which often use generalized linear models to quantify departures from neutral expectations[33,34,91]. However, these models generally assume that different lineages evolve independently and downplay the effect of genetic drift, thus potentially overestimating the lineage-specific changes due to positive selection[90]. Other methods incorporate phylogenetic relationships and account for genetic drift by modelling neutral evolution of expression levels through Brownian motion processes[61,88]. The joint action of genetic drift and stabilizing selection on gene expression can be modelled through Ornstein–Uhlenbeck processes[92,93], which allow both random variation and attraction towards an optimum value, for a quantitative trait[88,90]. These methods also test for the presence of lineage-specific optimum values of gene expression, which are potentially driven by positive selection, and a recent implementation accounts for environmental or technical variations among samples[39,90].

These methods have now been applied to both protein-coding gene expression evolution and alternative splicing frequencies[34], uncovering various lineage-specific expression changes that may be driven by positive selection. Thus, comparisons among primates revealed an over-representation of transcription factors among potentially selectively driven increases in gene expression in the human lineage[33,91], in agreement with the hypothesis that gene regulation differences may account for much of phenotypic variation[2]. Other significant expression shifts in the human lineage include genes involved in metabolic pathways, and this is consistent with the hypothesis that diet changes may be associated with human-specific adaptations[91,94] and with existing reports of positive selection in the promoters of nutrition-related genes[8]. The observed lineage-specific expression shifts are mostly organ specific[39,91]. In general, few lineage-specific expression shifts were observed in the brain, whereas the testis showed numerous expression shifts, consistent with the existence of different selective pressures on gene expression in these organs[39].

Challenges in the search for adaptive changes. The search for adaptive expression changes is challenging not least because of the existence of several confounding biological mechanisms. First, lineage-specific changes in cellular composition (such as white matter enrichment in the human prefrontal cortex[95]) are likely to be reflected in lineage-specific differences in expression levels when comparing whole organs, particularly for cell-type-specific genes[39]. Although these measured expression differences reflect marked phenotypic alterations and are thus biologically relevant, they do not represent actual cellular gene expression changes. Second, accelerated evolution may also be due to a non-adaptive mechanism: GC-biased gene conversion (gBGC), which

leads to preferential fixation of AT-to-GC mutations in regions with high recombination[96]. This mechanism can promote the fixation of slightly deleterious mutations and counterbalance the action of purifying selection[97]. Although the impact of gBGC on transcriptome evolution has yet to be assessed at a large scale, there is evidence that it is responsible for the human-specific accelerated evolution of a regulatory element that was originally attributed to positive selection[98,99]. Thus, gBGC should be considered when searching for adaptive transcriptome changes. Finally, changes in protein-coding transcription levels may be compensated by post-transcriptional or translational regulatory mechanisms that affect protein abundance or protein stability[100]. The presence of these additional layers of regulation relaxes the intensity of purifying selection at the transcriptional level, as gene expression changes at this level may have little or no phenotypic effect.

In addition to these biological factors, several technical limitations need to be taken into account. An important issue is the comparability of genomic annotations in gene expression analyses. Indeed, as alternatively spliced exons are pervasive in protein-coding genes[15], the inclusion of such exons in annotations of some but not all species may lead to apparent species- or lineage-specific expression shifts. Furthermore, as technical limitations of RNA-seq result in substantial read coverage variations along the gene, with a higher representation towards the 3′ end of the gene[101], uneven annotation in untranslated regions among species may also mimic expression changes. Thus, using comparable genomic annotations is paramount for identifying true lineage-specific expression changes[39].

Gene expression modules and networks. Traditionally, the search for an evolutionary connection between transcriptome and phenotype has relied on identifying expression changes of individual genes[33,34,39,91]. However, a more powerful approach may be the joint analysis of gene modules (that is, groups of genes with coherent expression patterns across a subset of samples), which could potentially include different combinations of species, organs and conditions[102]. This approach has revealed not only numerous modules of genes with highly conserved organ- or tissue-specific expression[39,103,104] but also simultaneous lineage-specific expression shifts[39,105]. For example, the module in FIG. 5a comprises 31 protein-coding genes enriched in nervous system development and neuron differentiation processes, and shows cerebellum-specific expression patterns in all analysed species[39]. However, cerebellum expression levels are higher in primates, suggesting a potentially coordinated lineage-specific expression change. This illustrates how gene module analyses can uncover potential lineage-specific expression shifts, even though they are not designed specifically for such a purpose.

In a related approach, expression patterns are analysed across multiple conditions (such as tissues, developmental stages and treatments) and species to reconstruct an evolutionarily conserved co-expression network[106]. Two genes are connected in this network if they show correlated expression levels in multiple species (FIG. 5). Co-expression connections are enriched for functionally related genes[106] and can thus predict functions for previously uncharacterized genes, including lncRNAs[51]. This approach led to the identification of a potential functional 'partner' for the *H19* lncRNA (FIG. 5b–d): *H19X* (also known as *MIR503HG*), a placenta-enriched X-linked lncRNA that acts as a miRNA precursor[51]. The expression pattern of *H19X* has strikingly shifted during evolution from ancestral testis specificity to placenta-predominant expression in placental mammals[51]. Thus, conserved co-expression networks may also reveal interesting transcriptome changes.

Parallel genome and transcriptome evolution. The search for adaptive transcriptome changes raises an important question: is expression evolution uncoupled from gene product (protein or RNA) evolution, or do the same sets of genes undergo adaptive changes at both levels? Several lines of evidence indicate that expression and gene product evolution generally occur in a parallel manner. First, genes specifically transcribed in the brain (the organ with the lowest rate of expression evolution) show slow protein sequence evolution. Conversely, genes predominantly expressed in the testis (the organ with the highest rate of expression evolution) have fast protein sequence evolution[10,39]. These observations confirm that the expression pattern is a strong determinant of the selective pressure acting on functional protein-coding regions[107]. Second, spatial expression pattern divergence and protein divergence are positively correlated regardless of the evolutionary distance between species[63]. Third, different gene categories show coordinated evolution in terms of expression pattern and gene products: protein-coding genes and miRNAs are generally highly conserved in both respects, whereas lncRNA sequences and expression patterns evolve rapidly[39,46,50,51]. However, beyond these general patterns, different gene categories may favour one of the two modes of evolution: genes with developmental and neural functions are skewed towards rapid expression evolution[63,108], whereas genes involved in immunity, male reproduction and olfaction are skewed towards rapid protein evolution[108].

Regulatory basis of transcriptome evolution

Numerous insights into the underlying biological mechanisms for transcriptome evolution have recently emerged from intraspecific and interspecific analyses of genome, epigenome and transcriptome variation. In particular, genome-wide association studies in human populations have identified DNA sequence variants that are potentially causative for various molecular phenotypes, including gene expression, alternative splicing patterns, histone modifications or transcription factor binding patterns[13,109].

Interspecies comparative analyses have also revealed mechanisms underlying transcriptome variation and their potential evolutionary importance. These include comparisons of genomic sequences that investigated

GC-biased gene conversion (gBGC). A non-adaptive evolutionary process that favours the fixation of GC alleles over AT alleles in highly recombining regions; it is likely to result from biased mismatch repair in heterozygous individuals during meiotic recombination.

the evolutionary forces acting on known regulatory elements. For example, lineage-specific changes in promoter sequences were reported to have been driven by positive selection — for example, opioid hormones[110] or genes involved in neural development[8] — although this finding may be confounded by an overall accelerated rate of neutral substitutions at human promoter regions[111]. Conserved non-coding elements (many of which may have regulatory functions) were also reported to show more evidence of adaptive evolution than protein-coding sequences in primates[112] and rodents[113]. However, in humans and chimpanzees most regulatory sequence divergence is likely to be due to accumulation of slightly deleterious mutations rather than to adaptive evolution[56,114], and gBGC has also resulted in rapid evolution of regulatory elements in humans[99].

With the advent of recent technological developments, gene expression regulatory mechanisms can now be more directly assessed. For example, cross-species chromatin immunoprecipitation followed by sequencing (ChIP–seq) analyses have revealed that binding of the liver-specific transcription factors CCAAT/enhancer-binding protein-α (CEBPA) and hepatocyte nuclear factor 4α (HNF4A) is highly species

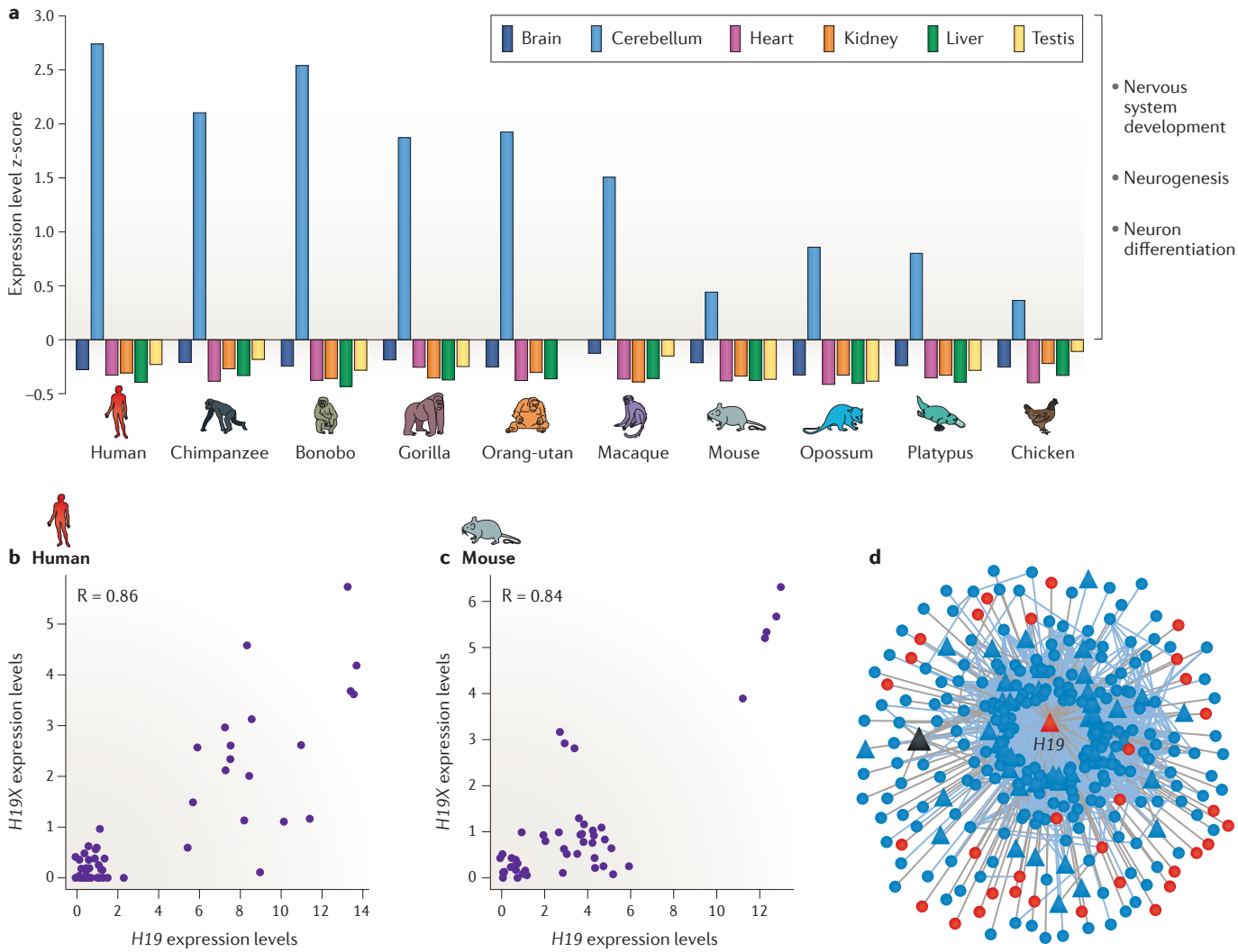


Figure 5 | **Expression modules and co-expression networks.** **a** | An example of an expression module (that is, a group of genes that show correlated expression patterns across a subset of samples) with conserved cerebellum-specific gene expression[39], which displays a potential increase in expression levels in the primate lineage, is shown. The genes in this module are enriched in gene ontology categories related to nervous system development. The bars represent the z-scores of gene expression levels, which were computed as the difference between the mean expression level of each organ and the average across all organs, divided by the standard deviation across all organs. High z-score values indicate high organ specificity. The z-scores were averaged across all 31 protein-coding genes present in the module. **b,c** | Correlation between the expression levels of the long non-coding RNAs (lncRNAs) *H19* and *H19X* in humans (part **b**) and mice (part **c**) is shown. The correlation coefficient (R) is also shown. **d** | Conserved co-expression network associated with the *H19* lncRNA is shown. Two genes are connected in the conserved co-expression network if their expression patterns are strongly correlated (either positively or negatively) in multiple species, as exemplified in parts **b** and **c** by *H19* and *H19X* in humans and mice. Blue points represent protein-coding genes, and red points represent lncRNAs. The blue triangles represent potential miRNA precursors, and the black triangle depicts *H19X*. Data in parts **b–d** were taken from REF. 51.

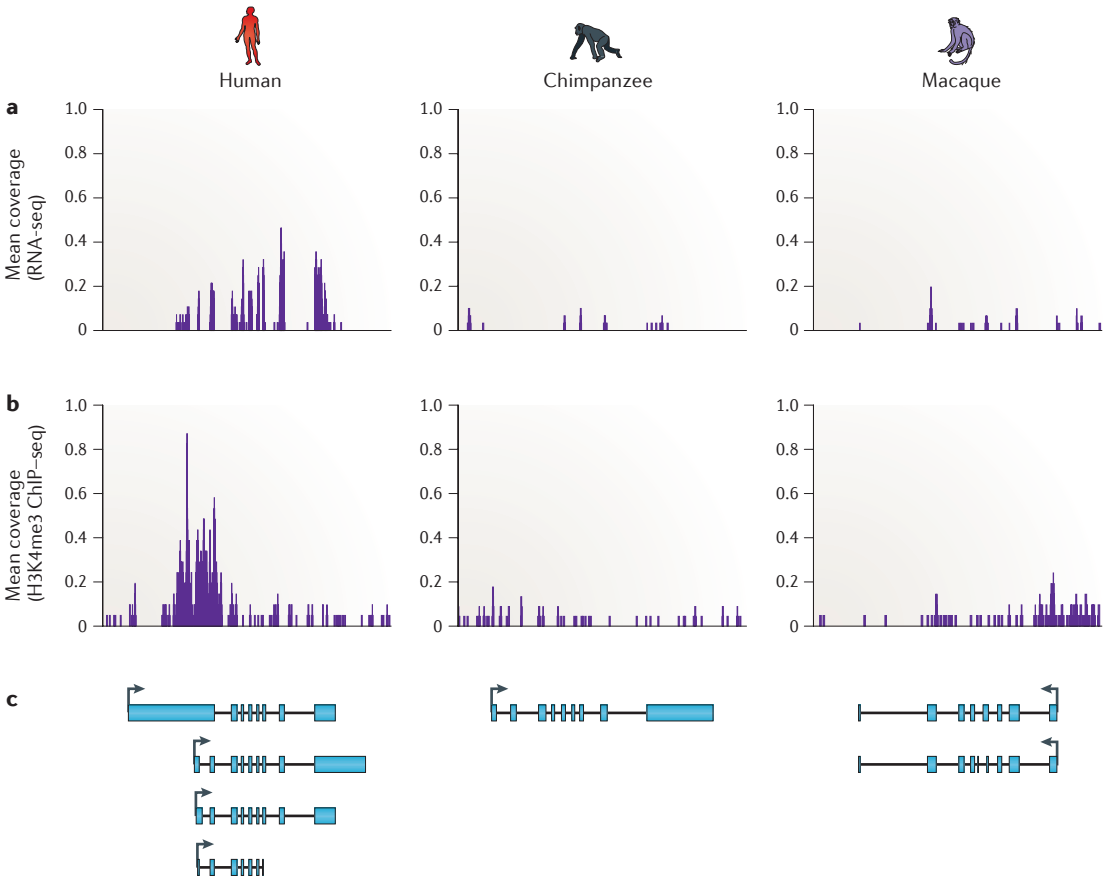


Figure 6 | **Example of a gene expression change associated with a change in histone methylation.** **a** | Variation of the mean RNA sequencing (RNA-seq) read coverage along the length of the *ELF3* gene in human, chimpanzee and macaque lymphoblastoid cell lines[120] is shown. **b** | Variation of the mean read coverage of chromatin immunoprecipitation followed by sequencing (ChIP–seq) of histone H3 lysine 4 trimethylation (H3K4me3) along the length of *ELF3* in human, chimpanzee and macaque lymphoblastoid cell lines[120] is shown. In parts **a** and **b**, the mean read coverage was normalized in each species by dividing it by the number of millions of mapped reads. **c** | Transcript annotations for *ELF3* in the same three species are shown. The arrows indicate the position of the annotated transcription start site and the direction of transcription. Rectangles indicate exon coordinates. Data were taken from REF. 120 and from the Ensembl database (release 75)[147].

specific[115], in contrast to the strong evolutionary conservation observed for the more general transcription factor CCCTC-binding factor (CTCF)[116]. Comparative analyses of DNA methylation[117,118] and histone modification patterns[119] in primates also identified numerous lineage-specific changes of these epigenetic marks, although overall chromatin profiles were similar in these species.

Remarkably, until now few studies have attempted to jointly analyse transcriptome evolution and underlying changes in regulation at a genome-wide scale. Among these, an analysis of histone H3 lysine 4 trimethylation (H3K4me3) patterns in lymphoblastoid cells estimated that as much as 7% of expression differences between humans, chimpanzees and macaques may be explained by differences in H3K4me3 status[120] (FIG. 6). Similarly, it was reported that differences in promoter CpG methylation may underlie up to 18% of gene expression changes between humans and chimpanzees in the liver, heart and kidney[118]. These studies indicate that epigenetic changes in *cis* might provide an explanation for a large proportion of gene expression evolution, although the causal relationship between these epigenetic features and gene expression patterns needs to be confirmed. Changes in the expression levels of *trans*-acting regulators (such as miRNAs) may also explain a non-negligible part (up to 6%) of transcriptome and proteome divergence between humans and chimpanzees in the adult brain[121]. Changes in miRNA expression are also associated with evolutionary divergence in developmental profiles of gene expression in the primate brain[64,122]. Moreover, the emergence of lineage-specific *trans*-acting factors may also have an important impact on transcriptome evolution, as illustrated by the human-specific miRNA miR-941, which affects genes involved in neurotransmitter signalling[123]. In general, the presence of lineage-specific miRNA genes with prevalent brain expression may have contributed to the 'rewiring' of regulatory networks in the mammalian brain[46].

Conclusions and future work

In the past few years, the development of new molecular technologies (particularly RNA-seq) has advanced the field of evolutionary transcriptomics. Thus, numerous comparative analyses have now been carried out for different aspects of the transcriptome, across multiple cell types, organs and developmental stages, for both closely related and distantly related species. These studies have revealed striking differences in evolutionary patterns among different parts of the transcriptome (including protein-coding and non-coding gene expression levels, as well as alternative splicing of protein-coding genes), mammalian lineages, chromosome types, cell types and organs. However, several aspects of mammalian transcriptome evolution have yet to be investigated in detail. For example, little is known about the evolution of untranslated regions of protein-coding genes, piRNAs or the recently discovered circular non-coding RNAs[124,125]. Moreover, with a few notable exceptions[64,122,126,127], most comparative transcriptomic studies so far have focused on adult organs, although most of the phenotypically relevant expression patterns may occur during embryonic development. The use of whole organs for evolutionary transcriptomics, which is predominant in existing studies, is also problematic owing to the potential presence of interspecies differences in cellular composition. Further technological innovations, such as the development of sensitive single-cell RNA-seq methods[128], may solve this limitation in the future.

Considerable advances were also recorded in the study of the underlying mechanisms of transcriptome changes. However, only few attempts have been made to study regulatory mechanisms and transcriptome patterns simultaneously. In particular, it is important to assess how changes in transcription factor binding patterns affect expression evolution, as it was recently reported that most interactions between transcription factors and chromatin do not result in significant changes in the expression of putative target genes[129]. Such studies would need to tackle the complex task of identifying accurate regulator–target gene relationships, which have so far often relied on the simplifying assumption that the closest gene to a transcription factor binding site or to an epigenetically marked region must be the target of regulation[130]. Future studies may benefit from the use of chromatin conformation capture techniques to identify distant regulatory elements[131].

Furthermore, an important limitation of current transcriptome evolution studies, particularly in the search for adaptive expression changes, is the assumption that changes in steady-state RNA levels (for example, as measured with RNA-seq) generally have a direct influence on the phenotype. However, interspecies differences in protein abundance are lower than those observed for mRNA levels, which indicates the presence of regulatory mechanisms that buffer transcription-level changes[100].

Finally, comparative transcriptomic studies have now identified numerous lineage-specific changes in expression levels, alternative splicing patterns or protein abundance levels in various lineages[34,39,100], which may contribute to the evolution of mammalian phenotypes. The next step is to carry out detailed functional characterizations of these molecular innovations in order to prove the gene expression–phenotype association. Pioneering studies have demonstrated that such work is feasible and rewarding in various model organisms (reviewed in REF. 12) not only for characterizing changes specific to these species (for example, through knockout or transgene experiments) but also for scrutinizing primate-specific alterations (for example, through *in vivo* replacements of mouse sequences with human variants).

1. Lynch, V. J., May, G. & Wagner, G. P. Regulatory evolution through divergence of a phosphoswitch in the transcription factor CEBPB. *Nature* **480**, 383–386 (2011).
2. King, M. C. & Wilson, A. C. Evolution at two levels in humans and chimpanzees. *Science* **188**, 107–116 (1975).
This early paper popularized the idea that the pronounced phenotypic differences between humans and chimpanzees may, to a large extent, be due to gene regulatory changes.
3. Britten, R. J. & Davidson, E. H. Gene regulation for higher cells: a theory. *Science* **165**, 349–357 (1969).
4. Hoekstra, H. E. & Coyne, J. A. The locus of evolution: evo devo and the genetics of adaptation. *Evolution* **61**, 995–1016 (2007).
5. Carroll, S. B. Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. *Cell* **134**, 25–36 (2008).
6. Wray, G. A. The evolutionary significance of *cis*-regulatory mutations. *Nature Rev. Genet.* **8**, 206–216 (2007).
7. Ponting, C. P. The functional repertoires of metazoan genomes. *Nature Rev. Genet.* **9**, 689–698 (2008).
8. Haygood, R., Fedrigo, O., Hanson, B., Yokoyama, K. D. & Wray, G. A. Promoter regions of many neural- and nutrition-related genes have experienced positive selection during human evolution. *Nature Genet.* **39**, 1140–1144 (2007).
9. Enard, W. *et al.* Intra- and interspecific variation in primate gene expression patterns. *Science* **296**, 340–343 (2002).
10. Khaitovich, P. *et al.* Parallel patterns of evolution in the genomes and transcriptomes of humans and chimpanzees. *Science* **309**, 1850–1854 (2005).
This study is one of the first to compare the patterns of protein-coding sequence evolution with the patterns of gene expression evolution in multiple human and chimpanzee organs.
11. Hawkins, R. D., Hon, G. C. & Ren, B. Next-generation genomics: an integrative approach. *Nature Rev. Genet.* **11**, 476–486 (2010).
12. Romero, I. G., Ruvinsky, I. & Gilad, Y. Comparative studies of gene expression and the evolution of gene regulation. *Nature Rev. Genet.* **13**, 505–516 (2012).
13. Lappalainen, T. & Dermitzakis, E. T. Evolutionary history of regulatory variation in human populations. *Hum. Mol. Genet.* **19**, R197–R203 (2010).
14. Nica, A. & Dermitzakis, E. T. Expression quantitative trait loci: present and future. *Phil. Trans. R. Soc. B* **368**, 20120362 (2013).
15. Wang, E. T. *et al.* Alternative isoform regulation in human tissue transcriptomes. *Nature* **456**, 470–476 (2008).
16. FANTOM Consortium and the RIKEN PMI and CLST (DGT) *et al.* A promoter-level mammalian expression atlas. *Nature* **507**, 462–470 (2014).
17. Ozsolak, F. *et al.* Comprehensive polyadenylation site maps in yeast and human reveal pervasive alternative polyadenylation. *Cell* **143**, 1018–1029 (2010).
18. Huntzinger, E. & Izaurralde, E. Gene silencing by microRNAs: contributions of translational repression and mRNA decay. *Nature Rev. Genet.* **12**, 99–110 (2011).
19. Ross, R. J., Weiner, M. M. & Lin, H. PIWI proteins and PIWI-interacting RNAs in the soma. *Nature* **505**, 353–359 (2014).
20. Brannan, C. I., Dees, E. C., Ingram, R. S. & Tilghman, S. M. The product of the H19 gene may function as an RNA. *Mol. Cell. Biol.* **10**, 28–36 (1990).
21. Guttman, M. *et al.* Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature* **458**, 223–227 (2009).
22. Cabili, M. N. *et al.* Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev.* **25**, 1915–1927 (2011).
23. Derrien, T. *et al.* The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Res.* **22**, 1775–1789 (2012).
24. Brown, C. J. *et al.* A gene from the region of the human X inactivation centre is expressed exclusively from the inactive X chromosome. *Nature* **349**, 38–44 (1991).
25. Latos, P. A. *et al. Airn* transcriptional overlap, but not its lncRNA products, induces imprinted *Igf2r* silencing. *Science* **338**, 1469–1472 (2012).
26. Orom, U. A. *et al.* Long noncoding RNAs with enhancer-like function in human cells. *Cell* **143**, 46–58 (2010).
27. Rinn, J. L. *et al.* Functional demarcation of active and silent chromatin domains in human *HOX* loci by noncoding RNAs. *Cell* **129**, 1311–1323 (2007).
28. Cesana, M. *et al.* A long noncoding RNA controls muscle differentiation by functioning as a competing endogenous RNA. *Cell* **147**, 358–369 (2011).

29. Keniry, A. *et al.* The *H19* lincRNA is a developmental reservoir of miR-675 that suppresses growth and *Igf1r*. *Nature Cell Biol.* **14**, 659–665 (2012).
30. Kim, T. K. *et al.* Widespread transcription at neuronal activity-regulated enhancers. *Nature* **465**, 182–187 (2010).
31. Marques, A. C. *et al.* Chromatin signatures at transcriptional start sites separate two equally populated yet distinct classes of intergenic long noncoding RNAs. *Genome Biol.* **14**, R131 (2013).
32. Duret, L., Chureau, C., Samain, S., Weissenbach, J. & Avner, P. The *Xist* RNA gene evolved in eutherians by pseudogenization of a protein-coding gene. *Science* **312**, 1653–1655 (2006).
33. 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).
34. Blekhman, R., Marioni, J. C., Zumbo, P., Stephens, M. & Gilad, Y. Sex-specific and lineage-specific alternative splicing in primates. *Genome Res.* **20**, 180–189 (2010).
35. 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).
36. Liao, B. Y. & Zhang, J. Evolutionary conservation of expression profiles between human and mouse orthologous genes. *Mol. Biol. Evol.* **23**, 530–540 (2006).
37. Chan, E. T. *et al.* Conservation of core gene expression in vertebrate tissues. *J. Biol.* **8**, 33 (2009).
38. Jordan, I. K., Marino-Ramirez, L. & Koonin, E. V. Evolutionary significance of gene expression divergence. *Gene* **345**, 119–126 (2005).
39. Brawand, D. *et al.* The evolution of gene expression levels in mammalian organs. *Nature* **478**, 343–348 (2011).
This is a comprehensive study of protein-coding gene expression level evolution based on RNA-seq transcriptome profiles for 6 major organs across 10 amniote species.
40. Modrek, B. & Lee, C. J. Alternative splicing in the human, mouse and rat genomes is associated with an increased frequency of exon creation and/or loss. *Nature Genet.* **34**, 177–180 (2003).
41. Pan, Q. *et al.* Alternative splicing of conserved exons is frequently species-specific in human and mouse. *Trends Genet.* **21**, 73–77 (2005).
42. Calarco, J. A. *et al.* Global analysis of alternative splicing differences between humans and chimpanzees. *Genes Dev.* **21**, 2963–2975 (2007).
43. Merkin, J., Russell, C., Chen, P. & Burge, C. B. Evolutionary dynamics of gene and isoform regulation in mammalian tissues. *Science* **338**, 1593–1599 (2012).
44. Barbosa-Morais, N. L. *et al.* The evolutionary landscape of alternative splicing in vertebrate species. *Science* **338**, 1587–1593 (2012).
References 43 and 44 present the first large-scale evolutionary comparison of alternative splicing patterns, which reveals rapid evolution of exon skipping frequencies.
45. Derti, A. *et al.* A quantitative atlas of polyadenylation in five mammals. *Genome Res.* **22**, 1173–1183 (2012).
46. Meunier, J. *et al.* Birth and expression evolution of mammalian microRNA genes. *Genome Res.* **23**, 34–45 (2013).
47. Sorefan, K. *et al.* Reducing ligation bias of small RNAs in libraries for next generation sequencing. *Silence* **3**, 4 (2012).
48. Raabe, C. A., Tang, T. H., Brosius, J. & Rozhdestvensky, T. S. Biases in small RNA deep sequencing data. *Nucleic Acids Res.* **42**, 1414–1426 (2014).
49. Roux, J., Gonzalez-Porta, M. & Robinson-Rechavi, M. Comparative analysis of human and mouse expression data illuminates tissue-specific evolutionary patterns of miRNAs. *Nucleic Acids Res.* **40**, 5890–5900 (2012).
50. Kutter, C. *et al.* Rapid turnover of long noncoding RNAs and the evolution of gene expression. *PLoS Genet.* **8**, e1002841 (2012).
51. Necsulea, A. *et al.* The evolution of lncRNA repertoires and expression patterns in tetrapods. *Nature* **505**, 635–640 (2014).
This is one of the first large-scale assessments of the evolutionary patterns of lncRNA repertoires and expression patterns across 11 tetrapod species.
52. Washietl, S., Kellis, M. & Garber, M. Evolutionary dynamics and tissue specificity of human long noncoding RNAs in six mammals. *Genome Res.* **24**, 616–628 (2014).
53. Chodroff, R. A. *et al.* Long noncoding RNA genes: conservation of sequence and brain expression among diverse amniotes. *Genome Biol.* **11**, R72 (2010).
54. Kutter, C. *et al.* Pol III binding in six mammals shows conservation among amino acid isotypes despite divergence among tRNA genes. *Nature Genet.* **43**, 948–955 (2011).
55. Li, W. H., Ellsworth, D. L., Krushkal, J., Chang, B. H. & Hewett-Emmett, D. Rates of nucleotide substitution in primates and rodents and the generation-time effect hypothesis. *Mol. Phylogenet Evol.* **5**, 182–187 (1996).
56. Keightley, P. D., Lercher, M. J. & Eyre-Walker, A. Evidence for widespread degradation of gene control regions in hominid genomes. *PLoS Biol.* **3**, e42 (2005).
57. Kaessmann, H., Heissig, F., von Haeseler, A. & Paabo, S. DNA sequence variation in a non-coding region of low recombination on the human X chromosome. *Nature Genet.* **22**, 78–81 (1999).
58. Yu, N., Fu, Y. X. & Li, W. H. DNA polymorphism in a worldwide sample of human X chromosomes. *Mol. Biol. Evol.* **19**, 2131–2141 (2002).
59. Kaessmann, H., Wiebe, V. & Paabo, S. Extensive nuclear DNA sequence diversity among chimpanzees. *Science* **286**, 1159–1162 (1999).
60. Warren, W. C. *et al.* Genome analysis of the platypus reveals unique signatures of evolution. *Nature* **453**, 175–183 (2008).
61. Khaitovich, P. *et al.* A neutral model of transcriptome evolution. *PLoS Biol.* **2**, E132 (2004).
62. Khaitovich, P., Enard, W., Lachmann, M. & Paabo, S. Evolution of primate gene expression. *Nature Rev. Genet.* **7**, 693–702 (2006).
63. Warnefors, M. & Kaessmann, H. Evolution of the correlation between expression divergence and protein divergence in mammals. *Genome Biol. Evol.* **5**, 1324–1335 (2013).
64. Somel, M. *et al.* MicroRNA-driven developmental remodeling in the brain distinguishes humans from other primates. *PLoS Biol.* **9**, e1001214 (2011).
This study presents a comprehensive analysis of miRNA and mRNA expression patterns across multiple developmental stages, brain regions and primate species, suggesting that differential regulation by miRNAs may underlie the human-specific developmental profiles.
65. Nielsen, R. *et al.* A scan for positively selected genes in the genomes of humans and chimpanzees. *PLoS Biol.* **3**, e170 (2005).
66. Soumillon, M. *et al.* Cellular source and mechanisms of high transcriptome complexity in the mammalian testis. *Cell Rep.* **3**, 2179–2190 (2013).
This study represents the first in-depth characterization of the transcriptome complexity of mammalian organs and reveals the cells and mechanisms underlying the high complexity of the testis transcriptome.
67. Kaessmann, H. Origins, evolution, and phenotypic impact of new genes. *Genome Res.* **20**, 1313–1326 (2010).
68. Kaessmann, H., Vinckenbosch, N. & Long, M. RNA-based gene duplication: mechanistic and evolutionary insights. *Nature Rev. Genet.* **10**, 19–31 (2009).
69. Giger, T. *et al.* Evolution of neuronal and endothelial transcriptomes in primates. *Genome Biol. Evol.* **2**, 284–292 (2010).
70. Yang, X. *et al.* Tissue-specific expression and regulation of sexually dimorphic genes in mice. *Genome Res.* **16**, 995–1004 (2006).
71. Reinius, B. *et al.* An evolutionarily conserved sexual signature in the primate brain. *PLoS Genet.* **4**, e1000100 (2008).
72. Veyrunes, F. *et al.* Bird-like sex chromosomes of platypus imply recent origin of mammal sex chromosomes. *Genome Res.* **18**, 965–973 (2008).
73. Cortez, D. *et al.* Origins and functional evolution of Y chromosomes across mammals. *Nature* **508**, 488–493 (2014).
This paper reports the first comprehensive analysis of sex chromosome origins and the functional evolution of Y-linked genes based on high-throughput genome and transcriptome sequencing data from representatives of all major mammalian lineages.
74. Bellott, D. W. *et al.* Mammalian Y chromosomes retain widely expressed dosage-sensitive regulators. *Nature* **508**, 494–499 (2014).
75. Charlesworth, B. Model for evolution of Y chromosomes and dosage compensation. *Proc. Natl Acad. Sci. USA* **75**, 5618–5622 (1978).
76. Ohno, S. *Sex Chromosomes and Sex-Linked Genes* (Springer-Verlag, 1967).
77. Julien, P. *et al.* Mechanisms and evolutionary patterns of mammalian and avian dosage compensation. *PLoS Biol.* **10**, e1001328 (2012).
This paper provides an initial overview of the patterns, mechanisms and evolution of X chromosome dosage compensation across mammals.
78. Lin, F., Xing, K., Zhang, J. & He, X. Expression reduction in mammalian X chromosome evolution refutes Ohno's hypothesis of dosage compensation. *Proc. Natl Acad. Sci. USA* **109**, 11752–11757 (2012).
79. Deng, X., Berletch, J. B., Nguyen, D. K. & Disteche, C. M. X chromosome regulation: diverse patterns in development, tissues and disease. *Nature Rev. Genet.* **15**, 367–378 (2014).
80. Zhang, Y. E., Vibranovski, M. D., Landback, P., Marais, G. A. & Long, M. Chromosomal redistribution of male-biased genes in mammalian evolution with two bursts of gene gain on the X chromosome. *PLoS Biol.* **8**, e1000494 (2010).
81. Rice, W. Sex chromosomes and the evolution of sexual dimorphism. *Evolution* **38**, 735–742 (1984).
82. Turner, J. M. Meiotic sex chromosome inactivation. *Development* **134**, 1823–1831 (2007).
83. Mueller, J. L. *et al.* The mouse X chromosome is enriched for multicopy testis genes showing postmeiotic expression. *Nature Genet.* **40**, 794–799 (2008).
84. Abzhanov, A., Protas, M., Grant, B. R., Grant, P. R. & Tabin, C. J. Bmp4 and morphological variation of beaks in Darwin's finches. *Science* **305**, 1462–1465 (2004).
85. Herrera, A. M., Shuster, S. G., Perriton, C. L. & Cohn, M. J. Developmental basis of phallus reduction during bird evolution. *Curr. Biol.* **23**, 1065–1074 (2013).
86. Manceau, M., Domingues, V.S., Mallarino, R. & Hoekstra, H.E. The developmental role of Agouti in color pattern evolution. *Science* **331**, 1062–1065 (2011).
87. Guenther, C. A., Tasic, B., Luo, L., Bedell, M. A. & Kingsley, D. M. A molecular basis for classic blond hair color in Europeans. *Nature Genet.* **46**, 748–752 (2014).
88. Bedford, T. & Hartl, D. L. Optimization of gene expression by natural selection. *Proc. Natl Acad. Sci. USA* **106**, 1133–1138 (2009).
89. Warnefors, M. & Eyre-Walker, A. A selection index for gene expression evolution and its application to the divergence between humans and chimpanzees. *PLoS ONE* **7**, e34935 (2012).
90. Rohlfs, R. V., Harrigan, P. & Nielsen, R. Modeling gene expression evolution with an extended Ornstein–Uhlenbeck process accounting for within-species variation. *Mol. Biol. Evol.* **31**, 201–211 (2014).
91. Blekhman, R., Oshlack, A., Chabot, A. E., Smyth, G. K. & Gilad, Y. Gene regulation in primates evolves under tissue-specific selection pressures. *PLoS Genet.* **4**, e1000271 (2008).
92. Hansen, T. F. Stabilizing selection and the comparative analysis of adaptation. *Evolution* **51**, 1341–1351 (1997).
93. Butler, M. A. & King, A. A. Phylogenetic comparative analysis: a modelling approach for adaptive evolution. *Am. Nat.* **164**, 683–695 (2004).
94. Babbitt, C. C., Warner, L. R., Fedrigo, O., Wall, C. E. & Wray, G. A. Genomic signatures of diet-related shifts during human origins. *Proc. Biol. Sci.* **278**, 961–969 (2011).
95. Schoenemann, P. T., Sheehan, M. J. & Glotzer, L. D. Prefrontal white matter volume is disproportionately larger in humans than in other primates. *Nature Neurosci.* **8**, 242–252 (2005).
96. Galtier, N. & Duret, L. Adaptation or biased gene conversion? Extending the null hypothesis of molecular evolution. *Trends Genet.* **23**, 273–277 (2007).
97. Galtier, N., Duret, L., Glemin, S. & Ranwez, V. GC-biased gene conversion promotes the fixation of deleterious amino acid changes in primates. *Trends Genet.* **25**, 1–5 (2009).

98. Prabhakar, S. *et al.* Human-specific gain of function in a developmental enhancer. *Science* **321**, 1346–1350 (2008).
99. Duret, L. & Galtier, N. Comment on "Human-specific gain of function in a developmental enhancer". *Science* **323**, 714 (2009).
100. Khan, Z. *et al.* Primate transcript and protein expression levels evolve under compensatory selection pressures. *Science* **342**, 1100–1104 (2013).
This study presents the first large-scale evidence that protein abundance levels evolve under stronger selective constraint than mRNA levels, indicating the presence of compensatory regulatory mechanisms.
101. Wang, Z., Gerstein, M. & Snyder, M. RNA-Seq: a revolutionary tool for transcriptomics. *Nature Rev. Genet.* **10**, 57–63 (2009).
102. Bergmann, S., Ihmels, J. & Barkai, N. Iterative signature algorithm for the analysis of large-scale gene expression data. *Phys. Rev. E Stat. Nonlin Soft Matter Phys.* **67**, 031902 (2003).
103. Piasecka, B., Kutalik, Z., Roux, J., Bergmann, S. & Robinson-Rechavi, M. Comparative modular analysis of gene expression in vertebrate organs. *BMC Genomics* **13**, 124 (2012).
104. Oldham, M.C., Horvath, S. & Geschwind, D.H. Conservation and evolution of gene coexpression networks in human and chimpanzee brains. *Proc. Natl Acad. Sci. USA* **103**, 17973–17978 (2006).
105. Konopka, G. *et al.* Human-specific transcriptional networks in the brain. *Neuron* **75**, 601–617 (2012).
106. Stuart, J. M., Segal, E., Koller, D. & Kim, S. K. A gene-coexpression network for global discovery of conserved genetic modules. *Science* **302**, 249–255 (2003).
107. Duret, L. & Mouchiroud, D. Determinants of substitution rates in mammalian genes: expression pattern affects selection intensity but not mutation rate. *Mol. Biol. Evol.* **17**, 68–74 (2000).
108. Haygood, R., Babbitt, C. C., Fedrigo, O. & Wray, G. A. Contrasts between adaptive coding and noncoding changes during human evolution. *Proc. Natl Acad. Sci. USA* **107**, 7853–7857 (2010).
109. Bryois, J. *et al. Cis* and *trans* effects of human genomic variants on gene expression. *PLoS Genet.* **10**, e1004461 (2014).
110. Rockman, M. V. *et al.* Ancient and recent positive selection transformed opioid *cis*-regulation in humans. *PLoS Biol.* **3**, e387 (2005).
111. Taylor, M. S. *et al.* Rapidly evolving human promoter regions. *Nature Genet.* **40**, 1262–1263 (2008).
112. Torgerson, D. G. *et al.* Evolutionary processes acting on candidate *cis*-regulatory regions in humans inferred from patterns of polymorphism and divergence. *PLoS Genet.* **5**, e1000592 (2009).
113. Halligan, D. L. *et al.* Contributions of protein-coding and regulatory change to adaptive molecular evolution in murid rodents. *PLoS Genet.* **9**, e1003995 (2013).
114. Gaffney, D. J., Blekhman, R. & Majewski, J. Selective constraints in experimentally defined primate regulatory regions. *PLoS Genet.* **4**, e1000157 (2008).
115. Schmidt, D. *et al.* Five-vertebrate ChIP–seq reveals the evolutionary dynamics of transcription factor binding. *Science* **328**, 1036–1040 (2010).
This paper presents the first large-scale evolutionary analysis of transcription factor binding patterns in five vertebrate species, which reveals species-specific binding patterns for two transcription factors.
116. Schmidt, D. *et al.* Waves of retrotransposon expansion remodel genome organization and CTCF binding in multiple mammalian lineages. *Cell* **148**, 335–348 (2012).
117. Hernando-Herraez, I. *et al.* Dynamics of DNA methylation in recent human and great ape evolution. *PLoS Genet.* **9**, e1003763 (2013).
118. Pai, A. A., Bell, J. T., Marioni, J. C., Pritchard, J. K. & Gilad, Y. A. Genome-wide study of DNA methylation patterns and gene expression levels in multiple human and chimpanzee tissues. *PLoS Genet.* **7**, e1001316 (2011).
119. Shulha, H. P. *et al.* Human-specific histone methylation signatures at transcription start sites in prefrontal neurons. *PLoS Biol.* **10**, e1001427 (2012).
120. Cain, C. E., Blekhman, R., Marioni, J. C. & Gilad, Y. Gene expression differences among primates are associated with changes in a histone epigenetic modification. *Genetics* **187**, 1225–1234 (2011).
121. Hu, H. Y. *et al.* MicroRNA expression and regulation in human, chimpanzee, and macaque brains. *PLoS Genet.* **7**, e1002327 (2011).
122. Somel, M. *et al.* MicroRNA, mRNA, and protein expression link development and aging in human and macaque brain. *Genome Res.* **20**, 1207–1218 (2010).
123. Hu, H. Y. *et al.* Evolution of the human-specific microRNA miR-941. *Nature Commun.* **3**, 1145 (2012).
124. Memczak, S. *et al.* Circular RNAs are a large class of animal RNAs with regulatory potency. *Nature* **495**, 333–338 (2013).
125. Hansen, T. B. *et al.* Natural RNA circles function as efficient microRNA sponges. *Nature* **495**, 384–388 (2013).
126. Somel, M. *et al.* Transcriptional neoteny in the human brain. *Proc. Natl Acad. Sci. USA* **106**, 5743–5748 (2009).
127. Irie, N. & Kuratani, S. Comparative transcriptome analysis reveals vertebrate phylotypic period during organogenesis. *Nature Commun.* **2**, 248 (2011).
128. Sandberg, R. Entering the era of single-cell transcriptomics in biology and medicine. *Nature Methods* **11**, 22–24 (2013).
129. Cusanovich, D. A., Pavlovic, B., Pritchard, J. K. & Gilad, Y. The functional consequences of variation in transcription factor binding. *PLoS Genet.* **10**, e1004226 (2014).
130. Gerstein, M. B. *et al.* Architecture of the human regulatory network derived from ENCODE data. *Nature* **489**, 91–100 (2012).
131. Dixon, J. R. *et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376–380 (2012).
132. Schena, M., Shalon, D., Davis, R. W. & Brown, P. O. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* **270**, 467–470 (1995).
133. Malone, J. H. & Oliver, B. Microarrays, deep sequencing and the true measure of the transcriptome. *BMC Biol.* **9**, 34 (2011).
134. Khaitovich, P. *et al.* Functionality of intergenic transcription: an evolutionary comparison. *PLoS Genet.* **2**, e171 (2006).
135. Oshlack, A., Chabot, A. E., Smyth, G. K. & Gilad, Y. Using DNA microarrays to study gene expression in closely related species. *Bioinformatics* **23**, 1235–1242 (2007).
136. Caceres, M. *et al.* Elevated gene expression levels distinguish human from non-human primate brains. *Proc. Natl Acad. Sci. USA* **100**, 13030–13035 (2003).
137. Fu, X. *et al.* Estimating accuracy of RNA-seq and microarrays with proteomics. *BMC Genomics* **10**, 161 (2009).
138. Garber, M., Grabherr, M. G., Guttman, M. & Trapnell, C. Computational methods for transcriptome annotation and quantification using RNA-seq. *Nature Methods* **8**, 469–477 (2011).
139. Perry, G. H. *et al.* Comparative RNA sequencing reveals substantial genetic variation in endangered primates. *Genome Res.* **22**, 602–610 (2011).
140. Dannemann, M. *et al.* Transcription factors are targeted by differentially expressed miRNAs in primates. *Genome Biol. Evol.* **4**, 552–564 (2012).
141. Li, Z. *et al.* Evolutionary and ontogenetic changes in RNA editing in human, chimpanzee, and macaque brains. *RNA* **19**, 1693–1702 (2013).
142. Warnefors, M., Liechti, A., Halbert, J., Valloton, D. & Kaessmann, H. Conserved microRNA editing in mammalian evolution, development and disease. *Genome Biol.* **15**, R83 (2014).
143. Lin, L. *et al.* Evolution of alternative splicing in primate brain transcriptomes. *Hum. Mol. Genet.* **19**, 2958–2973 (2010).
144. Babbitt, C. C. *et al.* Both noncoding and protein-coding RNAs contribute to gene expression evolution in the primate brain. *Genome Biol. Evol.* **2**, 67–79 (2010).
145. Yang, Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**, 1586–1591 (2007).
146. Vilella, A. J. *et al.* EnsemblCompara GeneTrees: complete, duplication-aware phylogenetic trees in vertebrates. *Genome Res.* **19**, 327–335 (2009).
147. Flicek, P. *et al.* Ensembl 2014. *Nucleic Acids Res.* **42**, D749–D755 (2014).

Acknowledgements
The authors apologize to colleagues whose work could not be discussed or cited owing to space constraints and/or the focus of this Review. This work was supported by grants from the European Research Council (Starting Grant: 242597, SexGenTransEvolution) and the Swiss National Science Foundation (Grant: 130287) to H.K. A.N. was supported by the Swiss National Science Foundation (Ambizione grant: 142636).

Competing interests statement
The authors declare no competing interests.