Published in final edited form as:
Nat Rev Genet. ; 13(7): 505–516. doi:10.1038/nrg3229.

Comparative studies of gene expression and the evolution of gene regulation

Irene Gallego Romero[1], **Ilya Ruvinsky**[2], and **Yoav Gilad**[1]

[1]Department of Human Genetics, University of Chicago, Chicago IL, USA

[2]Department of Ecology and Evolution, University of Chicago, Chicago IL, USA

Abstract

The hypothesis that differences in gene regulation play an important role in speciation and adaptation is more than 40 years old. With the advent of new sequencing technologies, we are able to characterize and study gene expression levels and associated regulatory mechanisms in a large number of individuals and species at unprecedented resolution and scale. We have thus gained new insights into the evolutionary pressures that shape gene expression levels, as well as developed an appreciation for the relative importance of evolutionary changes in different regulatory genetic and epigenetic mechanisms. The current challenge is to link gene regulatory changes to adaptive evolution of complex phenotypes. Here we mainly focus on comparative studies in primates, and how they are complemented by studies in model organisms.

Introduction

A major objective of evolutionary genetics is to provide a mechanistic account of the genetic basis for inter-species phenotypic variation. The goal is to identify the genetic changes and molecular mechanisms that underlie phenotypic diversity, as well as to understand the evolutionary pressures under which phenotypic diversity evolves. While the relative contribution of changes in gene regulation to adaptation continues to be debated[1, 2], it has become clear that variation in gene expression patterns often plays a key role in the evolution of morphological phenotypes[3] as well as a subset of other complex traits[4, 5].

The notion that changes in gene regulation often cause phenotypic diversity is not new. More than four decades ago, Britten and Davidson hypothesized in a series of papers[6, 7] that intergenic genomic regions (thought of by many at the time as ‘junk DNA’) play an important role in determining differences in gene regulatory patterns, and, consequently, phenotypic diversity. In 1975 King and Wilson[8] famously argued that the vast phenotypic differences between humans and chimpanzees are not likely to be explained solely by changes to structural proteins. They proposed that differences in gene regulation likely contribute to phenotypic differences between closely related species.

Corresponding author: Yoav Gilad (gilad@uchicago.edu) Dept of Human Genetics 920 E 58th Street Chicago, IL 60637 Phone: (773) 834-1984 Fax: (773) 834-8470.

For nearly 30 years, however, these hypotheses could not be rigorously tested or challenged, mainly because relevant data on gene regulation could not be collected at appropriate scale and resolution, and because of difficulties in identifying regulatory elements in the genome. It was also unclear to what extent the environment affects gene expression phenotypes, and whether it would at all be possible to detect genetic contributions to variation in gene regulation within or between species.

The last decade has seen tremendous developments in genomic technologies, which finally allowed investigators to apply high-throughput approaches to the study of gene expression patterns and associated regulatory mechanisms. For example, microarrays and now RNA sequencing (RNA-seq) enable genome-wide assessment of gene expression levels, and chromatin immunoprecipitation followed by sequencing (ChIP-seq) allows one to explore different aspects of regulatory mechanisms, such as transcription factor binding or histone modification. These advances provide the means to tackle outstanding questions regarding the evolution of gene regulation, including the characterization of the evolutionary forces that shape gene expression levels and the extent to which changes in different genetic and epigenetic mechanisms underlie regulatory variation. The relative importance of changes in gene regulation to phenotypic diversity and adaptation can now be studied with greater ease using these new techniques, although as we discuss below, a satisfying answer to this question still eludes us.

This review is focused on findings that emerge from comparative studies of gene regulation using cutting-edge genomic techniques. Studies that focus on variation in gene expression levels within species are discussed only briefly in this review. It is important to note, however, that the body of work focused on within-species patterns has provided important foundation for comparative studies by providing evidence that much of the observed variation in gene expression levels among individuals is heritable and can often be explained by corresponding genetic variation. Indeed it can often be mapped to specific loci referred to as expression quantitative trait loci, or eQTLs[9, 10]. This finding provided a strong motivation for comparative studies to focus on expression levels as an important intermediate molecular phenotype, one that ultimately determines heritable variation in complex morphological and physiological phenotypes, including traits that evolved under natural selection.

Early large-scale comparative studies of gene expression levels have been previously reviewed[11, 12]. Here we discuss recent progress in comparative studies of gene expression and regulation, primarily based on the use of new sequencing technologies. We start with an overview of comparative studies of gene expression levels and then explore observations – focusing on primates - that shed light on the evolution of gene regulation, and the associated genetic and epigenetic regulatory mechanisms. We discuss the connection between variation in gene regulation and variation in complex phenotypes, and in that context, point out important principal differences between comparative studies in primates and in model organisms. Finally, we comment on the possibilities to develop model systems that will allow us to further study the evolution of gene regulation in primates using experimental rather than strictly descriptive approaches.

Comparative studies of gene expression

A common approach to study of the evolution of gene regulation is to characterize and compare gene expression levels across species with the goal of understanding genetically regulated inter-species differences. Before the advent of next generation sequencing technologies, the only practical approach to measure and compare gene expression levels on a genome-wide scale was to use DNA microarrays. Comparative studies using arrays have resulted in important insight into the evolution of gene regulation (reviewed in[11-13]). Yet, microarrays can only be designed for species with available sequenced genomes. In contrast, using RNA sequencing (RNA-seq) techniques, one can measure and compare gene expression levels across practically any combination of species[14, 15], even when genomic sequences are not yet available[16]. In addition, RNA sequencing data allow one to estimate gene expression levels at a much broader dynamic range than microarrays, identify previously un-annotated transcripts, compare alternative splicing patterns and exon usage across species[17], and characterize genetic diversity in expressed genes[16]. Although comparative analysis of RNA sequencing data is challenging and remains an area of active research (Box 1), the advantages of this methodology over microarrays are clear[14, 18].

Inferring the action of natural selection on gene regulation

One approach to study the evolutionary forces that shape gene regulation is to identify gene expression patterns that can be explained by different evolutionary scenarios such as stabilizing or directional selection on gene regulation. To do so, one needs to distinguish between the environmental and genetic effects on gene regulation as well as control for a large number of potential sources of variation and error. These can be technical sources, such as variation in sample quality and batch effects (e.g., due to differences in collection protocols), or biological, such as variation due to sex, age, and circadian rhythm. In addition, physiological, morphological, and environmental differences between species (e.g., differences in diets) are also expected to contribute to differences in gene expression levels across species.

Studies in model organisms typically match the environmental conditions across individuals and take measures to minimize or control the technical and biological variation associated with the experiment. Comparative studies in model species can obtain evidence for natural selection on quantitative traits (such as gene expression levels) by testing for deviations from specified null models[19-21] (Box 2). Broadly speaking, this approach requires estimates of the expected inter-species variation in gene expression levels under the null (for instance, under a model of no selection), deviation from which are interpreted as evidence for alternative scenarios (for example, evidence for the action of natural selection). Such an approach relies on a number of parameter estimates (for example, the mutation accumulation rate), which need to be estimated or measured independently[13].

In non-model organisms, notably in primates, it is often impossible or impractical to directly estimate the parameters of a null model of the evolution of gene expression. One alternative to specifying an explicit model is to take an empirical approach, in which genes are first ranked according to their patterns of expression levels within and between species, and then evaluated for fit to expectations under different evolutionary scenarios (Box 2). The goal of

the empirical approach is to identify specific patterns of heritable gene expression levels, which are consistent with the action of natural selection. However, in non-model organisms it is often impossible to distinguish between technical and biological variance or to match the environment across individuals of different species. As a result, some observations from comparative studies of gene regulation in such species should be interpreted with caution.

The observation of inter-species differences in gene expression levels is inherently difficult to interpret, because environmental and genetic explanations can be completely confounded. It is reasonable to assume that differences in environment experienced by different individuals and species will generally result in perturbation of gene regulation and lead to an increase in variation of gene expression levels. In contrast, genes that have low variation in expression levels across individuals and species are probably those that are robust to environmental differences. One can therefore conclude with considerable confidence that the regulation of genes with constant expression levels across individuals and species is genetically controlled. Low variation in gene expression levels across species is consistent with the action of stabilizing selection on gene regulation[24]. When a difference in gene expression is seen in a specific lineage (Box 2) - for example, a higher expression level observed exclusively in humans - this may indicate the action of directional selection on gene regulation in that lineage. Alternatively, it may be a consequence of a specific environmental influence on that lineage (for example, the consumption of cooked food in the case of humans[22, 23]).

Comparative studies of gene expression in primates

Differences in gene regulation between humans and other primates may ultimately be used to explain the molecular basis for human-specific traits. For example, it was hypothesized that human-specific gene expression patterns in the brain[25, 26] might underlie functional, developmental, and perhaps cognitive differences between humans and other apes. A recent comparative study that incorporated temporal resolution into the study design found potential differences in the timing of gene expression in the brain across primates[27], which might be related to inter-species differences in timing of developmental processes. Genes with potential roles in neural development showed a marked delay in expression timing in human brain samples compared with chimpanzee and rhesus macaque[27]. More generally, several major principles have emerged from comparative studies of gene expression among primates (and in some cases among other species as well).

Selective constraint

Although the notion that the expression levels of most genes are shaped by natural selection was once debated[28], multiple studies now support the conclusion that the regulation of a large subset of genes and pathways evolve under natural selection in primates[27, 29, 30]. Comparative gene expression data in apes and old world monkeys suggest that the regulation of a large subset of genes is evolving under selective constraint. Indeed, comparative studies[27, 29, 30] have found that the extent of inter-species variation in gene expression levels can often be explained by variation in gene expression within a species, consistent with the action of stabilizing selection on gene regulation. More generally, though there is much uncertainty about the relevant values of important parameters for a standard

neutral model of gene expression evolution in primates (as discussed above and in box 2), even when conservative estimates are used for generation time and mutation rates, the overwhelming majority of genes exhibit far less between species variation in gene expression levels than expected if all regulatory mutations were neutral[19]. These studies, however, had the minor weakness because they relied only on comparative data from closely related species (typically, humans, chimpanzees, and rhesus macaques). Thus it remained possible that the inference of widespread selective constraint on gene regulation could be explained by lack of mutations that effected gene expression due to chance. That is, because regulatory elements constitute a small fraction of the genome, gene expression patterns among closely related species may appear to be under constraint if not enough time has passed since the most recent common ancestor for regulatory substitutions to accumulate in substantial numbers.

More recently, an RNA-seq study has looked at gene expression levels and genetic diversity in livers from 16 mammalian species, including humans and 11 non-human primates[16]. All liver samples for this study were collected postmortem and it was therefore not possible to stage the tissues or control for possible environmental effects across species. Nevertheless, expression patterns of many genes showed remarkable conservation, suggesting a strong genetic component in their regulation as well as the action of stabilizing selection over hundreds of millions of years.

Directional selection

There is also evidence that the regulation of some genes - 10-30% of genes (depending on the tissue / cell type studied)[31-33] - has evolved under directional (positive) selection. For instance, the comparative RNA-seq study of 16 species[16] also identified lineage-specific changes in expression levels; an example is shown in Box 3. However, as we discussed above, inferring positive directional selection on gene regulation in non-model species is more complicated than inferring selective constraint. Although a lineage-specific change in gene expression level may be consistent with the action of directional selection - that is, it is reasonable to assume that directional selection on gene regulation would result in inter-species differences in gene expression levels - it is unclear how many regulatory differences are truly the result of selection. Alternative explanations for gene expression differences between species, such as consistent inter-species differences in environments, are often difficult to exclude, especially in primates. By ranking genes according to inter-individual variation in expression levels one can confidently assume that the set of genes that are differentially expressed among species and are associated with low within-species variance – as a group - is enriched for targets of selection compared to genes that are not differentially expressed between species (Box 2). Yet, it may always be difficult to identify with confidence the individual genes whose regulation evolved under positive selection.

Tissue-specificity

Another question is whether gene regulation in primates evolves under tissue-specific selection pressures. A recent RNA-seq study[15] estimated gene expression levels in six different tissues from nine mammalian species (including humans and all four great apes using data from in) and showed significantly different rates of transcriptome evolution

across tissues. This study[15] identified 145 gene expression network modules that had lineage-specific expression patterns, which may indicate the action of species-specific and tissue-specific directional selection on gene regulation. This study also found 33 organ-specific gene expression network modules that are conserved across these mammals and are enriched with genes involved in biological processes intuitively considered typical for each of the studied tissues (e.g. synaptic transmission in the brain). Similar patterns were observed in a more limited comparative study in humans, chimpanzees and rhesus macaques, which focused on gene expression measurements from hearts, livers, and kidneys from multiple individuals[33]. In the most extreme cases, the observed inter-species expression patterns of a subset of genes were consistent with the action of stabilizing selection in one tissue (e.g., liver), and the action of lineage-specific directional selection in another tissue (e.g. heart). The results of these studies are consistent with the idea that adaptation may more commonly proceed via regulatory rather than structural (i.e. coding) changes, because regulatory mutations have spatially or temporally circumscribed effects.

Alternative splicing

The third emerging principle is that inter-species differences in gene expression levels only rarely can be explained by differences in alternative splicing between species. It may seem surprising, because alternative splicing and changes in exon usage could provide an intuitive mechanism with which to introduce functional variation to structural proteins. Yet, only few instances of inter-species differences in exon usage have been observed[15, 16, 30]. For example, a recent study sequenced liver RNA from males and females of humans, chimpanzees, and rhesus macaques and characterized gene and exon-specific expression levels. This study showed that while sexually dimorphic differences in exon usage are relatively common, sexually dimorphic gene expression levels and alternative splicing patterns are largely conserved between species[30]. A caveat of this result is that non-human primate transcriptomes are not well annotated, so that the probability of missing an exon expressed only in a non-human primate may be high. However, such technical explanations are unlikely to account for the observation that nearly all expressed exons in humans are also expressed in non-human primates. Given the sequencing depth of recent comparative studies, explanations based on lack of power are unlikely either.

As can be seen, comparative studies in primates, while challenging, have resulted in important insights into the evolution of gene expression levels. Yet, we are also finding that gene expression patterns alone provide little insight into the adaptive phenotypes, molecular mechanisms, or even the specific biological processes involved in the observed changes in gene expression levels. The question at this point is how to move beyond descriptive studies of gene expression levels across species?

From gene expression to regulatory mechanisms

There are two general approaches to ‘move beyond’ a simple description of the evolution of gene expression patterns. One is to perform functional experiments to understand adaptive phenotypes; the question that is typically being asked is “what differences in phenotype do these changes in gene expression levels underlie?” The other general approach is to perform comparative studies of the underlying regulatory mechanisms; in effect pursuing the

opposite direction, as it were, asking “what changes in regulatory mechanisms explain the observed differences in gene expression levels?” The latter approach does not provide insight into phenotypes, but it addresses other outstanding questions regarding the mechanisms that s hape regulatory evolution (Figure 1). In this section we discuss the progress that has been made using comparative studies of regulatory mechanisms.

A large number of gene regulatory mechanisms are reasonably well understood (for example those involved in transcription initiation; reviewed in [34]). Yet, we still know little about the relative contribution of changes in different genetic and epigenetic regulatory mechanisms to the evolution of gene expression levels. From an evolutionary biologists' perspective, uncovering the mechanisms of regulatory adaptations will reveal what types of mutations underlie inter-species differences in gene expression levels and reveal the genetic loci that likely underlie phenotypic adaptation and speciation. From a biomedical perspective, understanding mechanisms of regulatory evolution, especially in primates, is expected to help us guide the search for functional elements in the human genome, which are likely to disproportionally harbor disease-causing mutations[35].

Comparative studies of regulatory mechanisms need to address the same challenges and difficulties that were discussed in the context of comparative gene expression studies. Genetic and epigenetic regulatory profiles are influenced by environment, cell composition, and circadian rhythm, just to name a few potentially confounding effects. It is easier to control for these effects when conducting studies in model organisms but, nevertheless, important trends have emerged from comparative studies in primates as well.

Comparisons of transcription factor binding

In one of the first sequence-based comparative functional genome-wide studies of transcription factor binding[36], ChIP-seq was performed for two hepatic transcription factors in liver samples from five vertebrates, including humans. The results showed that most binding locations are species-specific. Of the ~16,000-30,000 binding sites identified in each species, only 35 were shared across all five species, and only 344 were shared by the three mammalian species studied (humans, mice and dogs). A study of RNA Pol II binding[37] showed that 32% of binding locations in immortalized B cell lines differed between humans and chimpanzee (although it is important to note that they only had one chimpanzee sample), and 25% of sites differed between human individuals. These studies suggest that evolutionary turnover of transcription factor binding sites is rapid and that, on a genome-wide scale, most binding locations may not be conserved even across closely related species (Figure 2). However, because these studies did not collect comparative gene expression data from the same samples, it was not possible to assess the degree to which differences in transcription factor binding might account for inter-species differences in gene expression levels. As a result, it cannot be excluded that those binding events that have effects on gene regulation are more conserved than suggested by general genome-wide patterns.

A different approach was taken in a study[38] that introduced a functional and freely segregating copy of human chromosome 21 into a mouse to generate a model of trisomy 21. Examination of the binding locations of three transcription factors - HNF1a, HNF4a, and HNF - in livers from these mice and in human hepatocytes showed that 85-92% of binding

locations on human chromosome 21 in the mouse coincided with binding sites observed in normal human hepatocytes[39]. Moreover, the expression profiles of genes on human chromosome 21 in mouse hepatocytes were highly correlated with those from human hepatocytes. Thus, in this case, differences in the cellular environment between human and mouse livers resulted in relatively little change in transcription factor binding or gene expression patterns. The important inference from this study is that the sequence of human chromosome 21 appears to encode sufficient information to result in faithful regulatory output in mouse, namely, regardless of the cellular environment.

Comparisons of chromatin state and DNA methylation

Another trend that emerges from comparative studies of regulatory mechanisms, especially in primates, is that a substantial fraction of gene expression differences across species can be explained by inter-species changes in epigenetic mechanisms. For instance, genomic regions associated with H3K4me3 - a histone mark that denotes active transcription[40] – were characterize using ChIP-seq in immortalized B cells from humans, chimpanzees, and rhesus macaques[41] and RNA-seq data were also collected from the same samples. Overall, there were large differences in the patterns of this histone modification across the three species, but a high degree of conservation near transcription start sites (TSS), where H3K4me3 is most likely to be functional. The subset of genes associated with inter-species differences in H3K4me3 modification near their TSS were also more likely to be differentially expressed between species. Because this study looked at correlations between gene expression data and H3K4me3 ChIP-seq data, direct causal inference was impossible. Nevertheless, based on previous work on regulation by histone modifications[42, 43] the authors estimated that up to 7% of gene expression differences across the three species could be accounted for by changes in H3K4me3 status.

A similar approach was used to study correlations between gene expression levels and promoter DNA methylation status in livers, hearts, and kidneys from humans and chimpanzees[44]. As expected, variation in methylation states between different tissues was greater than between species. Moreover, tissue-specific promoter methylation profiles were generally conserved. This result is consistent with other studies that reported a large overlap in methylation profiles across primates - for example, in human and chimpanzee sperm[45], or in human, chimpanzee and orangutan neutrophils[46]. That said, differentially expressed genes between humans and chimpanzees were often associated with promoter methylation differences, regardless of tissue. Based on a large body of work that supports the causal effects of promoter DNA methylation on gene regulation[47, 48], the authors estimated that as much as 12-18% (depending on the tissue) of inter-species differences in gene expression levels could be explained by changes in promoter methylation profiles.

As these examples illustrate, most comparative work to date has focused on mechanisms of transcriptional initiation. A few studies, however, are looking elsewhere for factors that can influence gene regulation during evolution. For instance, changes in microRNA expression levels, which are expected to affect rates of mRNA decay, could account for ~2-4% of gene expression differences across the prefrontal cortex of humans, chimpanzees and rhesus macaques[49, 50].

From gene expression to complex phenotypes

Comparative studies of regulatory mechanisms in primates rely on correlations between different measurements. Despite important insights, without direct experimentation it is difficult to assess causality or the impact of changes in regulatory mechanisms on gene expression levels at the organism level. Functional experimentation in humans and other apes is technically limited to a few immortalized cell lines, non-invasively sampled tissues, or post-mortem samples, which are difficult to stage. In most cases it is difficult to infer which phenotypic adaptation was mediated by species-specific changes in gene expression levels or even how to formulate specific hypotheses for further experiments. Even when the mechanism and specific regulatory sequence elements underlying the expression change may be known (e.g., using the approaches described above to characterize the regulatory mechanisms), the phenotypes that are being affected by the regulatory change are typically unknown. Because of the obvious ethical and practical limitations on experimentation in primates (especially apes), it is difficult to envision an approach that will allow one to follow-up these observations and test their functional relevance. To circumvent these limitations, several studies have utilized model organisms to address specific hypotheses inspired by comparative analysis of gene regulation in primates.

For example, McLean and colleagues[51] investigated the phenotype associated with a human-specific 5 kb deletion upstream of the androgen receptor (*AR*) gene, which include sequence that is conserved in other mammals (and therefore is likely to be functional). Constructs containing the mouse and chimpanzee versions of this region directed reporter gene expression in the facial vibrissae and genital tubercle of transgenic mice. Since *AR* is implicated in the development of sensory vibrissae and penile spines[52, 53], the loss of this tissue-specific enhancer in the human lineage was interpreted as a causal mechanism for the human-specific loss of these morphological properties.

Other studies, using similar approaches that involve functional experimentation in model systems, identified an ancient enhancer that may have recently gained a human-specific function linked with the evolution of the human thumb[43], a change in non-coding RNA sequence that may be linked to cortical development[54], and a human-specific change in the forkhead transcription factor *FOXP2*, which might be related to the development of language[55, 56]. It should be noted, however, that in most of these studies model organisms are used to recapitulate gene regulatory differences between primates and to study them with high spatial and temporal resolution[57]. Therefore, the inference about function requires one to make two important assumptions. First, that the effects of gene regulatory changes on complex phenotypes are identical in model organisms and in primates, including humans. This assumption may be difficult to accept in some cases, for example when the phenotype under consideration is language. Second, that no other regulatory changes could manifest in similar patterns. For example, if multiple enhancers drive nearly identical spatio-temporal expression patterns of a reporter gene, it is unclear how to identify the particular enhancer whose evolution may be associated with a derived trait. At the moment, data are not yet available to estimate how often this assumption is reasonable.

Comparative studies of gene regulation in model organisms

Because a broad range of experimental manipulations are possible in model organisms, studies that focus on model species can move beyond simple comparisons of gene expression and offer deep insights into the causal relationship between regulatory changes and phenotypic evolution. Consider, for example, a pair of species that are distinguished by a specific difference in morphology, physiology, or behavior. Such a difference might result in a fitness benefit in the environments which the species inhabit, thereby revealing a selective pressure under which it has evolved (demonstrating this is often quite challenging; see Barrett and Hoekstra[58] for a recent review). The two species may be sufficiently closely related to permit crosses, in which the genetic determinants of the inter-species phenotypic differences could be mapped. One can then use different techniques (e.g., positional cloning) to identify the specific mutations and molecular mechanisms underlying the phenotypic divergence and provide evidence for causality.

A compelling example is the case of pelvic fin reduction in a threespine stickleback (*Gasterosteus aculeatus*). Repeated instances of pelvic reduction are thought to be adaptive and associated with invasions into fresh water habitats. *Pitx1*, a gene encoding a transcription factor involved in pelvic fin development, has been identified as a candidate locus responsible for this morphological change[59]. Fine mapping[60] pointed to a putative regulatory element upstream of *Pitx1* as the causal locus. The deletion of this regulatory element, which population genetic data suggest has been subjected to positive selection, was hypothesized to result in a difference in *Pitx1* expression pattern and, ultimately, in a reduced pelvic fin. This hypothesis was supported by transgenic experiments that demonstrated that the candidate noncoding region is indeed a regulatory enhancer. Furthermore, the reduced-pelvic phenotype could be reversed by using a transgene containing the candidate genomic region. Similarly compelling examples are the change in a *cis*-regulation of the *Agouti* gene during the evolution of camouflage coloration in *Peromyscus* mice[61] and the regulatory change of the *optix* gene, which has been identified as the site of repeated evolution of the wing color patterns responsible for mimicry in *Heliconius* butterflies[62].

More generally, work in model species suggest that divergence of gene expression levels of individual loci may be subtle[63, 64], but that even small changes in regulatory state can cause substantial phenotypic divergence[65] associated with fitness effects[66, 67]. This view emphasizes the complex polygenic nature of the evolution of gene expression[68], one in which epistatic interactions[69, 70] and interactions with the environment[71] are important. That said, studies in which both the evolutionary history and the molecular mechanisms are well understood remain relatively rare. In contrast, quite a few studies in model organisms have identified clear connections between changes in gene regulation and differences in phenotypes, which are assumed to be adaptive. While the plausible scenario of adaptation can often be proposed, the exact nature of it remains elusive. Examples from plants, fungi, and animals illustrate the breadth of this phenomenon (e.g., [69, 72-75]). Of course, evolution of many traits is not caused by changes in gene regulation[76, 77]. Dramatic examples include a single amino acid mutation in the *melanocortin-1 receptor* gene causing pigmentation

differences in beach mice[78] and *aquaporin* gene loss in natural populations of *S. cerevisiae*[79].

Emerging principles from studies of model organisms

In contrast to studies in primates, studies in model organisms have resulted in much more direct insight into the mechanisms underlying the evolution of gene regulation. For instance, it is difficult to obtain data that conclusively supports regulatory changes in *cis* or *trans* in primates (the regulatory inferences we discussed above cannot easily be validated or confirmed), but this has been done many times in model systems. Changes in *cis* elements appear to be more commonly responsible for inter-species differences in gene expression patterns than changes in *trans*, as shown in yeast and flies[80-83]. One mechanism that can lead to *cis*-regulatory divergence is a rapid turnover of transcription factor binding sites[84], which in turn could cause different transcription factor binding profiles, even between closely related species[85]. Changes in trans-regulatory elements (such as transcription factors and regulatory RNAs) have also been documented in yeast[82, 86], and there is considerable evidence of co-evolution of *cis* and *trans* regulatory elements in various species[82, 87, 88].

In addition, several lines of evidence implicate chromatin state as an important player in the evolution of gene expression. Circumstantial evidence for the importance of this mechanism comes from studies in primates as well[89, 90], but in model systems it is possible to directly demonstrate causality. Studies in yeast have shown that despite an overall similarity in nucleosome positioning profiles, genes with divergent expression often show divergent chromatin organization[91-94]. Furthermore, certain properties of nucleotide sequences predispose promoters to evolve divergent gene expression more readily, perhaps via changes in chromatin structure[95]. For example, deletions of chromatin factors in yeast revealed previously cryptic gene expression differences, suggesting that these proteins buffer regulatory variation[96].

Recent experimental results in model systems[74, 97-99] are also resurrecting the classical idea that transposable elements, containing pre-existing transcription factor binding sites, could insert in the vicinity of regulatory loci, and serve as a source of novel regulatory elements[6]. It appears that latent regulatory activity can be located in introns[75] and even deteriorating coding sequences[100]. Whereas most studies discussed here considered transcriptional gene regulation, many other molecular processes regulate gene expression and can thus contribute to evolution of gene expression and phenotypes[101, 102].

Conclusions

Genomic technologies allow us to characterize variation in gene expression levels within and between species with relative ease. As might be expected, the data suggest that the regulation of most genes evolved under evolutionary constraint, though subsets of genes whose regulation likely has evolved under directional selection can also be found. The challenge is to move beyond comparative descriptions of gene expression levels to the study of the underlying mechanisms and the connection between regulatory evolution and ultimate adaptation of complex phenotypes.

The lofty promise of genomics – to predict functional elements, including regulatory loci, based on primary sequence – is becoming a reality. Major advances have been made in developing quantitative predictions of gene expression patterns based on *cis*-regulatory sequences[103]. At first these models primarily considered interactions of transcription factors with DNA[104, 105], but more recently they have started to incorporate nucleosome-positioning information[106, 107], making predictions more accurate and biologically realistic. Much work is still required, but as more sophisticated models are developed, we will likely improve on our current ability to predict gene expression patterns from the sequences of their regulatory elements[108]. This, in turn, will help to determine which of the millions of nucleotide differences between the genomes of related species are responsible for their divergent patterns of gene regulation.

Functional studies of variation in complex phenotypes, however, will always be needed to validate model predictions, and these must involve empirical approaches. As we have discussed, although progress has been slow in all systems, effective experiments can be designed for model organisms. One can reveal the causal relationships between differences in gene expression levels, the underlying regulatory mechanisms, and the evolution of complex phenotypes. In primates, the only functional approach available thus far is to rely on experimentation on model systems, a useful approach at times, but the results of which are often somewhat difficult to interpret. If we are ever to utilize comparative functional approaches to study the genetic architecture that underlies regulatory adaptation and its phenotypic consequences in humans and other apes, a new paradigm is needed. Perhaps the advent of induced pluripotent stem cells (iPSCs) will provide an alternative system for functional studies in primates. iPSCs can be differentiated into a multitude of cell types, and thus provide a surrogate system in which to functionally test the links between inter-species changes in gene regulation and differences in phenotypes. Admittedly, even under the best-case scenario one could only focus on cellular phenotypes. Yet, the wide range of cell types that can potentially be derived from iPSCs (e.g., hepatocytes, cardiomyocytes, neurons) will offer a range of molecular phenotypes to choose from, perhaps finally making a reality detailed mechanistic functional studies of gene expression evolution in primates.

Acknowledgments

We thank M. Nobrega and N. Sakab for helping to generate figure 1, M. Ward and D. Odom for generating figure 2 based on their comparative data, and G. Perry for help with the figures in boxes 2 and 3. We thank J. Pritchard, Z. Gauhar and three anonymous reviewers for comments on the manuscript. This work was supported by NSF grant IOS-0843504 and NIH grant P50 GM081892 to IR, and NIH grants GM077959 and GM084996 to YG. IGR is a Sir Henry Wellcome Postdoctoral Fellow.

Biography

Yoav Gilad is an Associate Professor in the Department of Human Genetics at the University of Chicago. He studies the evolution of gene regulation in primates with the long-term goal of identifying the genetic basis for human specific traits, including genetic variation that underlies higher susceptibility to certain diseases and disorders in humans than in other primates. In addition to characterizing gene expression levels, the Gilad lab studies variation in regulatory mechanisms, in order to understand how genetic and epigenetic inter-species differences translate into gene regulatory differences between species. Most recently

the Gilad lab has turned to functional studies in cellular systems to investigate the phenotypic consequences of inter-species regulatory changes.

Irene Gallego Romero is a Sir Henry Wellcome Postdoctoral Fellow in the Gilad laboratory at the University of Chicago. She is interested in the role natural selection plays in shaping humans, both as a species and as distinct populations subject to particular selective pressures, as well as in identifying and unraveling the mechanisms through which regulatory changes give rise to these phenotypic differences. At present her work is focused on identifying difference in gene expression patterns during organ development between humans and non-human primates by using cellular systems.

Ilya Ruvinsky is an Associate Professor in the Department of Ecology and Evolution at the University of Chicago. A major goal of his laboratory is to elucidate molecular mechanisms and functional consequences of gene regulatory evolution. His research interests also include genome evolution, interactions between organisms and the environment, and systems biology. He received a Ph.D. in Molecular Biology from Princeton University and went on to conduct postdoctoral research in the Department of Molecular Biology, Massachusetts General Hospital and Department of Genetics, Harvard Medical School.

Glossary

Pelvic fin	The fins that are attached to the pelvic girdle, on the lower surface of the fish body. They help control the direction of movement.
Enhancer	A region of DNA that binds to proteins whose function is to promote transcription of genes.
Mimicry	When an organism benefits from copying the phenotype of another organism.
***trans*-regulatory elements**	Regulatory elements that can affect the transcription rates of both alleles of a gene (examples include transcription factors and small regulatory RNAs). In contrast, *cis*-regulatory elements have an allele-specific regulatory effect.
Transposable elements	DNA sequences that can change their position in the genome.
The neutral model	A model that states that alleles that reach sufficient frequency within a population to be sampled, or are fixed between species, are selectively neutral, whereas a subset of alleles are too strongly deleterious to either segregate within a population in appreciable frequencies or reach fixation.
Ranking-based approach	Genome-wide studies often use model-free ranking to prioritize candidate genes. Ranking is performed based on properties that are expected to be informative with respect to the desired trait (for example, nucleotide diversity across populations when the desired traits is evidence for natural selection).

Vitamin A toxicity	Having too much vitamin A in the body. This can lead to multiple clinically abnormal conditions including decreased appetite, softening of the skull bone, nausea, vomiting, blurry vision, headaches, and hair loss.
MNase sequencing	Sequencing of chromatin that has been treated with micrococcal nuclease (MNase), which preferentially cuts linker DNA connecting two nucleosomes. MNase sequencing can be used to map nucleosome positions.
RNA sequencing (RNA-seq)	An experimental protocol that uses next-generation sequencing technologies to sequence the RNA molecules within a biological sample in an effort to determine the primary sequence and relative abundance of each RNA type.
Expression QTL	(eQTL). A locus at which genetic allelic variation is associated with variation in gene expression levels.
Positional cloning	A method for identifying the location of a risk variant within a candidate region. Overlapping clones covering the candidate region are typed, and segments that co-segregate perfectly with the disease are identified. These clones are the most likely location of the risk variant.
Induced pluripotent stem cells	These are derived from somatic cells by ‘reprogramming’ or de-differentiation triggered by the transfection of pluripotency genes, which alters the somatic cells to a state that is similar to that of embryonic stem cells.
Stabilizing selection	Natural selection against individuals that deviate from an intermediate optimum; this process tends to stabilize the phenotype. By contrast, directional selection pushes it towards either extreme.

References

1. Carroll SB. Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. Cell. 2008; 134:25–36. [PubMed: 18614008]
2. Hoekstra HE, Coyne JA. The locus of evolution: evo devo and the genetics of adaptation. Evolution. 2007; 61:995–1016. [PubMed: 17492956] [References 1 and 2 summarize the on-going controversy regarding the relative importance of changes to structural proteins and changes in gene regulation to adaptation and speciation.]
3. Stern DL, Orgogozo V. The Loci of Evolution: How Predictable Is Genetic Evolution? Evolution. 2008:2155–2177. [PubMed: 18616572]
4. Kleinjan DA, van Heyningen V. Long-range control of gene expression: emerging mechanisms and disruption in disease. American Journal of Human Genetics. 2005; 76:8–32. [PubMed: 15549674]
5. Wray GA. The evolutionary significance of cis-regulatory mutations. Nature Reviews Genetics. 2007; 8:206–16.
6. Britten RJ, Davidson EH. Gene regulation for higher cells: a theory. Science. 1969; 165:349–57. [PubMed: 5789433]

7. Britten RJ, Davidson EH. Repetitive and non-repetitive DNA sequences and a specualation on the origins of evolutionary novelty. Quarterly Review of Biology. 1971; 46:111–138. [PubMed: 5160087]
8. King, M.-c.; Wilson, A. Evolution at Two Levels in Humans and Chimpanzees. Science. 1975; 188:107–116. [PubMed: 1090005]
9. Majewski J, Pastinen T. The study of eQTL variations by RNA-seq: from SNPs to phenotypes. Trends in Genetics. 2011; 27:72–9. [PubMed: 21122937]
10. Gilad Y, Rifkin SA, Pritchard JK. Revealing the architecture of gene regulation: the promise of eQTL studies. Trends Genet. 2008; 24:408–15. [PubMed: 18597885]
11. Zheng W, Gianoulis TA, Karczewski KJ, Zhao H, Snyder M. Regulatory variation within and between species. Annu Rev Genomics Hum Genet. 2011; 12:327–46. [PubMed: 21721942]
12. Whitehead A, Crawford DL. Neutral and adaptive variation in gene expression. Proc Natl Acad Sci U S A. 2006; 103:5425–30. [PubMed: 16567645]
13. Gilad Y, Oshlack A, Rifkin SA. Natural selection on gene expression. Trends Genet. 2006; 22:456–61. [PubMed: 16806568]
14. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. Nature Reviews Genetics. 2009; 10:57–63.
15. Brawand D, et al. The evolution of gene expression levels in mammalian organs. Nature. 2011; 478:343–348. [PubMed: 22012392] [To date, this is the largest and most comprehensive investigation of gene expression evolution across a wide range of vertebrate animals and in multiple tissues.]
16. Perry GH, et al. Comparative RNA sequencing reveals substantial genetic variation in endangered primates. Genome Res. 2012
17. Gelfman S, et al. Changes in exon-intron structure during vertebrate evolution affect the splicing pattern of exons. Genome Research. 2012; 22:35–50. [PubMed: 21974994]
18. Marioni JC, Mason CE, Mane SM, Stephens M, Gilad Y. RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. Genome Research. 2008; 18:1509–17. [PubMed: 18550803]
19. Lemos B, Meiklejohn CD, Caceres M, Hartl DL. Rates of divergence in gene expression profiles of primates, mice, and flies: stablizing selection and variability among functional categories. Evolution. 2005; 59:126–137. [PubMed: 15792233] [This study analyzed gene expression data from multiple species and concluded that the regulation of the vast majority of genes evolves under evolutionary constraint.]
20. Rifkin SA, Kim J, White KP. Evolution of gene expression in the Drosophila melanogaster subgroup. Nat Genet. 2003; 33:138–44. [PubMed: 12548287]
21. Rifkin SA, Houle D, Kim J, White KP. A mutation accumulation assay reveals extensive capacity for rapid gene expression evolution. Nature. 2005; 438:220–3. [PubMed: 16281035] [This was the first mutation accumulation study that estimated the neutral rate of gene expression divergence in drosophila. Based on their estimates the authors concluded that the regulation of most genes evolve under evolutionary constraint.]
22. Wrangham RW, Jones JH, Laden G, Pilbeam D, Conklin-Brittain N. The Raw and the Stolen. Cooking and the Ecology of Human Origins. Curr Anthropol. 1999; 40:567–594. [PubMed: 10539941]
23. Finch CE, Stanford CB. Meat-adaptive genes and the evolution of slower aging in humans. Quarterly Review of Biology. 2004; 51:3–50. [PubMed: 15101252]
24. Gilad Y, Oshlack A, Smyth GK, Speed TP, White KP. Expression profiling in primates reveals a rapid evolution of human transcription factors. Nature. 2006; 440:242–5. [PubMed: 16525476]
25. Nowick K, Gernat T, Almaas E, Stubbs L. Differences in human and chimpanzee gene expression patterns define an evolving network of transcription factors in brain. Proceedings of the National Academy of Sciences. 2009; 106:22358–63.
26. Guohua Xu A, et al. Intergenic and repeat transcription in human, chimpanzee and macaque brains measured by RNA-Seq. PLoS Computational Biology. 2010; 6:e1000843. [PubMed: 20617162]
27. Somel M, et al. Transcriptional neoteny in the human brain. Proceedings of the National Academy of Sciences. 2009; 106:5743–8.

28. Khaitovich P, et al. A neutral model of transcriptome evolution. PLoS Biology. 2004; 2:E132. [PubMed: 15138501]
29. Khaitovich P, Enard W, Lachmann M, Paabo S. Evolution of primate gene expression. Nat Rev Genet. 2006; 7:693–702. [PubMed: 16921347]
30. Blekhman R, Marioni JC, Zumbo P, Stephens M, Gilad Y. Sex-specific and lineage-specific alternative splicing in primates. Genome Research. 2010; 20:180–9. [PubMed: 20009012]
31. Enard W, et al. Intra- and interspecific variation in primate gene expression patterns. Science. 2002; 296:340–3. [PubMed: 11951044]
32. Gilad Y, Oshlack A, Smyth GK, Speed TP, White KP. Expression profiling in primates reveals a rapid evolution of human transcription factors. Nature. 2006; 440:242–5. [PubMed: 16525476]
33. Blekhman R, Oshlack A, Chabot AE, Smyth GK, Gilad Y. Gene regulation in primates evolves under tissue-specific selection pressures. PLoS Genet. 2008; 4:e1000271. [PubMed: 19023414]
34. Valen E, Sandelin A. Genomic and chromatin signals underlying transcription start-site selection. Trends in genetics. 2011; 27:475–485. [PubMed: 21924514]
35. Cooper GM, Shendure J. Needles in stacks of needles: finding disease-causal variants in a wealth of genomic data. Nat Rev Genet. 2011; 12:628–40. [PubMed: 21850043]
36. Schmidt D, et al. Five-vertebrate ChIP-seq reveals the evolutionary dynamics of transcription factor binding. Science. 2010; 328:1036–40. [PubMed: 20378774] [This was the first comparative ChIP-seq study in vertebrate. The results indicate extensive turnover of regulatory elements across even relatively closely related species.]
37. Kasowski M, et al. Variation in transcription factor binding among humans. Science. 2010; 328:232–5. [PubMed: 20299548]
38. O'Doherty A, et al. An aneuploid mouse strain carrying human chromosome 21 with Down syndrome phenotypes. Science. 2005; 309:2033–7. [PubMed: 16179473]
39. Wilson MD, et al. Species-specific transcription in mice carrying human chromosome 21. Science. 2008; 322:434–8. [PubMed: 18787134] [An elegant study design allowed the authors to demonstrate that regulatory sequences are largely sufficient to direct transcriptional programs, even when the cellular environment is changing.]
40. Heintzman ND, et al. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. Nature Genetics. 2007; 39:311–8. [PubMed: 17277777]
41. Cain CE, Blekhman R, Marioni JC, Gilad Y. Gene expression differences among primates are associated with changes in a histone epigenetic modification. Genetics. 2011; 187:1225–34. [PubMed: 21321133]
42. Schneider R, et al. Histone H3 lysine 4 methylation patterns in higher eukaryotic genes. Nature Cell Biology. 2004; 6:73–7.
43. Prabhakar S, et al. Human-specific gain of function in a developmental enhancer. Science. 2008; 321:1346–50. [PubMed: 18772437] [This was one of the first studies that used a model system to try to assign functional significance to observed regulatory evolution in primates. By expressing primate enhancers in mouse, the authors showed that a recently diverged human enhancer drives strong reporter gene expression in limb while the orthologous elements from chimpanzee and rhesus macaque do not.]
44. Pai AA, Bell JT, Marioni JC, Pritchard JK, Gilad Y. A genome-wide study of DNA methylation patterns and gene expression levels in multiple human and chimpanzee tissues. PLoS Genetics. 2011; 7:e1001316. [PubMed: 21383968]
45. Molaro A, et al. Sperm Methylation Profiles Reveal Features of Epigenetic Inheritance and Evolution in Primates. Cell. 2011; 146:1029–1041. [PubMed: 21925323]
46. Martin DI, et al. Phyloepigenomic comparison of great apes reveals a correlation between somatic and germline methylation states. Genome Res. 2011; 21:2049–57. [PubMed: 21908772]
47. Murrell A, Rakyan VK, Beck S. From genome to epigenome. Human Molecular Genetics. 2005; 14 Spec No 1:R3–R10. [PubMed: 15809270]
48. Jaenisch R, Bird A. Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. Nature Genetics. 2003; 33(Suppl):245–54. [PubMed: 12610534]
49. Hu HY, et al. MicroRNA Expression and Regulation in Human, Chimpanzee, and Macaque Brains. PLoS Genetics. 2011; 7:e1002327. [PubMed: 22022286]

50. Somel M, et al. MicroRNA-Driven Developmental Remodeling in the Brain Distinguishes Humans from Other Primates. PLoS Biology. 2011; 9:e1001214. [PubMed: 22162950]
51. McLean CY, et al. Human-specific loss of regulatory DNA and the evolution of human-specific traits. Nature. 2011; 471:216–9. [PubMed: 21390129]
52. Crocoll A, Zhu CQC, Cato ACB, Blum M. Expression of androgen receptor mRNA during mouse embryogenesis. Mechanisms of Development. 1998; 72:175–178. [PubMed: 9533962]
53. Dixson AF. Effects of testosterone on the sternal cutaneous glands and genitalia of the male greater galago (Galago crassicaudatus crassicaudatus). Folia Primatologica (Basel). 1976; 26:207–13.
54. Pollard KS, et al. An RNA gene expressed during cortical development evolved rapidly in humans. Nature. 2006; 443:167–72. [PubMed: 16915236]
55. Enard W, et al. A humanized version of Foxp2 affects cortico-basal ganglia circuits in mice. Cell. 2009; 137:961–71. [PubMed: 19490899]
56. Enard W, et al. Molecular evolution of FOXP2, a gene involved in speech and language. Nature. 2002; 418:869–72. [PubMed: 12192408]
57. Sholtis SJ, Noonan JP. Gene regulation and the origins of human biological uniqueness. Trends in Genetics. 2010
58. Barrett RD, Hoekstra HE. Molecular spandrels: tests of adaptation at the genetic level. Nat Rev Genet. 2011; 12:767–80. [PubMed: 22005986]
59. Shapiro MD, et al. Genetic and developmental basis of evolutionary pelvic reduction in threespine sticklebacks. Nature. 2004; 428:717–23. [PubMed: 15085123] [This was one of the first papers that started a detailed body of work by the same group on the regulatory basis for the evolution of pelvic reduction in sticklebacks. Work in sticklebacks (see also reference 60) has allowed investigators to provide a truly detailed account of the mechanistic basis for this morphological adaptation.]
60. Chan YF, et al. Adaptive Evolution of Pelvic Reduction in Sticklebacks by Recurrent Deletion of a Pitx1 Enhancer. Science. 2010; 327:302–305. [PubMed: 20007865]
61. Manceau M, Domingues VS, Mallarino R, Hoekstra HE. The developmental role of Agouti in color pattern evolution. Science. 2011; 331:1062–5. [PubMed: 21350176] [(along with Reference 78): In model systems it is possible to truly dissect the genetic basis for adaptations, as well as to test for different evolutionary scenarios. This paper and reference 78 do so with respect to the evolution of coat color in Peromyscus mice (see reference 78 as well).]
62. Reed RD, et al. optix drives the repeated convergent evolution of butterfly wing pattern mimicry. Science. 2011; 333:1137–41. [PubMed: 21778360]
63. Barriere A, Gordon KL, Ruvinsky I. Distinct Functional Constraints Partition Sequence Conservation in a cis-Regulatory Element. PLoS Genetics. 2011; 7
64. Crocker J, Tamori Y, Erives A. Evolution Acts on Enhancer Organization to Fine-Tune Gradient Threshold Readouts. PLoS Biology. 2008; 6:2576–2587.
65. Fowlkes CC, et al. A Conserved Developmental Patterning Network Produces Quantitatively Different Output in Multiple Species of Drosophila. PLoS Genetics. 2011; 7
66. Ludwig MZ, Manu Kittler, R. White KP, Kreitman M. Consequences of Eukaryotic Enhancer Architecture for Gene Expression Dynamics, Development, and Fitness. PLoS Genetics. 2011; 7
67. Ludwig MZ, et al. Functional evolution of a cis-regulatory module. PLoS Biology. 2005; 3:588–598.
68. Bullard JH, Mostovoy Y, Dudoit S, Brem RB. Polygenic and directional regulatory evolution across pathways in Saccharomyces. Proceedings of the National Academy of Sciences. 2010; 107:5058–5063.
69. Gerke J, Lorenz K, Cohen B. Genetic Interactions Between Transcription Factors Cause Natural Variation in Yeast. Science. 2009; 323:498–501. [PubMed: 19164747]
70. Gertz J, Gerke JP, Cohen BA. Epistasis in a quantitative trait captured by a molecular model of transcription factor interactions. Theoretical Population Biology. 2010; 77:1–5. [PubMed: 19818800]
71. Gerke J, Lorenz K, Ramnarine S, Cohen B. Gene-Environment Interactions at Nucleotide Resolution. PLoS Genetics. 2010; 6

72. Doebley J, Stec A, Gustus C. Teosinte Branched1 and the Origin of Maize - Evidence for Epistasis and the Evolution of Dominance. Genetics. 1995; 141:333–346. [PubMed: 8536981]
73. Clark RM, Wagler TN, Quijada P, Doebley J. A distant upstream enhancer at the maize domestication gene tb1 has pleiotropic effects on plant and inflorescent architecture. Nature Genetics. 2006; 38:594–597. [PubMed: 16642024]
74. Studer A, Zhao Q, Ross-Ibarra J, Doebley J. Identification of a functional transposon insertion in the maize domestication gene tb1. Nature Genetics. 2011; 43:1160–U164. [PubMed: 21946354]
75. Rebeiz M, Jikomes N, Kassner VA, Carroll SB. Evolutionary origin of a novel gene expression pattern through co-option of the latent activities of existing regulatory sequences. Proceedings of the National Academy of Sciences. 2011; 108:10036–10043.
76. Hoekstra HE, Coyne JA. The locus of evolution: evo devo and the genetics of adaptation. Evolution. 2007; 61:995–1016. [PubMed: 17492956]
77. Stern DL, Orgogozo V. The loci of evolution: how predictable is genetic evolution? Evolution. 2008; 62:2155–77. [PubMed: 18616572]
78. Hoekstra HE, Hirschmann RJ, Bundey RA, Insel PA, Crossland JP. A single amino acid mutation contributes to adaptive beach mouse color pattern. Science. 2006; 313:101–4. [PubMed: 16825572]
79. Will JL, et al. Incipient balancing selection through adaptive loss of aquaporins in natural Saccharomyces cerevisiae populations. PLoS Genet. 2010; 6:e1000893. [PubMed: 20369021]
80. Brem RB, Yvert G, Clinton R, Kruglyak L. Genetic dissection of transcriptional regulation in budding yeast. Science. 2002; 296:752–5. [PubMed: 11923494]
81. Ronald J, Brem RB, Whittle J, Kruglyak L. Local regulatory variation in Saccharomyces cerevisiae. PLoS Genetics. 2005; 1:e25. [PubMed: 16121257]
82. Tirosh I, Reikhav S, Levy AA, Barkai N. A Yeast Hybrid Provides Insight into the Evolution of Gene Expression Regulation. Science. 2009; 324:659–662. [PubMed: 19407207]
83. Wittkopp PJ, Haerum BK, Clark AG. Regulatory changes underlying expression differences within and between Drosophila species. Nature Genetics. 2008; 40:346–350. [PubMed: 18278046] [This study was the first to provide genome-wide estimates of the relative proportion of changes in cis and trans regulatory elements that underlie differences in gene expression levels within and between species.]
84. Doniger SW, Fay JC. Frequent gain and loss of functional transcription factor binding sites. PLoS Computational Biology. 2007; 3:932–942.
85. Bradley RK, et al. Binding Site Turnover Produces Pervasive Quantitative Changes in Transcription Factor Binding between Closely Related Drosophila Species. PLoS Biology. 2010; 8
86. Tirosh I, Sigal N, Barkai N. Divergence of nucleosome positioning between two closely related yeast species: genetic basis and functional consequences. Molecular Systems Biology. 2010; 6
87. Gordon KL, Ruvinsky I. Tempo and mode in evolution of transcriptional regulation. PLoS Genetics. 2012; 8:e1002432. [PubMed: 22291600]
88. Landry CR, et al. Compensatory cis-trans evolution and the dysregulation of gene expression in interspecific hybrids of Drosophila. Genetics. 2005; 171:1813–22. [PubMed: 16143608]
89. Ernst J, et al. Mapping and analysis of chromatin state dynamics in nine human cell types. Nature. 2011; 473:43–9. [PubMed: 21441907]
90. Degner JF, et al. DNase I sensitivity QTLs are a major determinant of human expression variation. Nature. 2012; 482:390–4. [PubMed: 22307276]
91. Field Y, et al. Gene expression divergence in yeast is coupled to evolution of DNA-encoded nucleosome organization. Nature Genetics. 2009; 41:438–445. [PubMed: 19252487] [This study is one of the first to demonstrate the role of changes in nucleosome positioning to the evolution of gene expression levels across species (see references 86 and 96 as well).]
92. Tsankov A, Yanagisawa Y, Rhind N, Regev A, Rando OJ. Evolutionary divergence of intrinsic and trans-regulated nucleosome positioning sequences reveals plastic rules for chromatin organization. Genome Research. 2011; 21:1851–1862. [PubMed: 21914852]
93. Tsankov AM, Thompson DA, Socha A, Regev A, Rando OJ. The Role of Nucleosome Positioning in the Evolution of Gene Regulation. PLoS Biology. 2010; 8

94. Tsui K, et al. Evolution of Nucleosome Occupancy: Conservation of Global Properties and Divergence of Gene-Specific Patterns. Molecular and Cellular Biology. 2011; 31:4348–4355. [PubMed: 21896781]
95. Vinces MD, Legendre M, Caldara M, Hagihara M, Verstrepen KJ. Unstable tandem repeats in promoters confer transcriptional evolvability. Science. 2009; 324:1213–6. [PubMed: 19478187]
96. Tirosh I, Reikhav S, Sigal N, Assia Y, Barkai N. Chromatin regulators as capacitors of interspecies variations in gene expression. Molecular Systems Biology. 2010; 6
97. Bourque G, et al. Evolution of the mammalian transcription factor binding repertoire via transposable elements. Genome Research. 2008; 18:1752–1762. [PubMed: 18682548]
98. Lynch VJ, Leclerc RD, May G, Wagner GP. Transposon-mediated rewiring of gene regulatory networks contributed to the evolution of pregnancy in mammals. Nature Genetics. 2011; 43:1154–U158. [PubMed: 21946353]
99. Smith AM, et al. A novel mode of enhancer evolution: The Tal1 stem cell enhancer recruited a MIR element to specifically boost its activity. Genome Research. 2008; 18:1422–1432. [PubMed: 18687876]
100. Eichenlaub MP, Ettwiller L. De Novo Genesis of Enhancers in Vertebrates. PLoS Biology. 2011; 9
101. Alonso CR, Wilkins AS. The molecular elements that underlie developmental evolution. Nature Reviews Genetics. 2005; 6:709–715.
102. Dori-Bachash M, Shema E, Tirosh I. Coupled Evolution of Transcription and mRNA Degradation. PLoS Biology. 2011; 9
103. Segal E, Widom J. From DNA sequence to transcriptional behaviour: a quantitative approach. Nature Reviews Genetics. 2009; 10:443–456.
104. Janssens H, et al. Quantitative and predictive model of transcriptional control of the Drosophila melanogaster even skipped gene. Nature Genetics. 2006; 38:1159–1165. [PubMed: 16980977]
105. Segal E, Raveh-Sadka T, Schroeder M, Unnerstall U, Gaul U. Predicting expression patterns from regulatory sequence in Drosophila segmentation. Nature. 2008; 451:535–U1. [PubMed: 18172436]
106. Raveh-Sadka T, Levo M, Segal E. Incorporating nucleosomes into thermodynamic models of transcription regulation. Genome Research. 2009; 19:1480–1496. [PubMed: 19451592]
107. Wasson T, Hartemink AJ. An ensemble model of competitive multi-factor binding of the genome. Genome Research. 2009; 19:2101–2112. [PubMed: 19720867]
108. Landolin JM, et al. Sequence features that drive human promoter function and tissue specificity. Genome Research. 2010; 20:890–898. [PubMed: 20501695]
109. Kimura M. Genetic variability maintained in a finite population due to mutational production of neutral and nearly neutral isoalleles. Genet Res. 1968; 11:247–69. [PubMed: 5713805]
110. Kimura, M. The neutral theory. Cambridge university press; Cambridge: 1983. p. 34-55.
111. Lande R. Natural-Selection and Random Genetic Drift in Phenotypic Evolution. Evolution. 1976; 30:314–334.
112. Lynch M, Hill WG. Phenotypic Evolution by Neutral Mutation. Evolution. 1986; 40:915–935.
113. Pickrell JK, et al. Signals of recent positive selection in a worldwide sample of human populations. Genome Research. 2009; 19:826–37. [PubMed: 19307593]
114. Sabeti PC, et al. Genome-wide detection and characterization of positive selection in human populations. Nature. 2007; 449:913–8. [PubMed: 17943131]
115. Voight BF, Kudaravalli S, Wen X, Pritchard JK. A map of recent positive selection in the human genome. PLoS Biology. 2006; 4:e72. [PubMed: 16494531]
116. Williamson SH, et al. Localizing recent adaptive evolution in the human genome. PLoS Genetics. 2007; 3:e90. [PubMed: 17542651]
117. Teshima KM, Coop G, Przeworski M. How reliable are empirical genomic scans for selective sweeps? Genome Res. 2006; 16:702–12. [PubMed: 16687733]
118. Akey JM. Constructing genomic maps of positive selection in humans: where do we go from here? Genome Research. 2009; 19:711–22. [PubMed: 19411596]

119. Tan CL, Drake JH. Evidence of tree gouging and exudate eating in pygmy slow lorises (Nycticebus pygmaeus). Folia Primatol (Basel). 2001; 72:37–9. [PubMed: 11275748]

120. Swapna N, Radhakrishna S, Gupta AK, Kumar A. Exudativory in the Bengal slow loris (Nycticebus bengalensis) in Trishna Wildlife Sanctuary, Tripura, northeast India. Am J Primatol. 2010; 72:113–21. [PubMed: 19937974]

121. Sakabe NJ, Nobrega MA. Genome-wide maps of transcription regulatory element. Wiley Interdiscip. Rev. Syst. Biol. Med. 2010; 2:422–437. [PubMed: 20836039]

122. Scally A, et al. Insights into hominid evolution from the gorilla genome sequence. Nature. 2012; 483:169–75. [PubMed: 22398555]

Box 1 – Comparative analysis of RNA sequencing data

Comparative studies of gene expression levels using RNA sequencing, overcomes many of the traditional limitations associated with microarray data, but are not free of challenges. Most challenges are common to all RNAseq studies and relate to the count nature of the RNA-seq data, the need to normalize and standardize the data, and the desire to account for confounding and biasing factors (such as differences in transcript length or GC content across genes). One challenge, however, is fairly specific to comparative studies: the requirement of defining the transcriptome. This is necessary because comparisons of expression level estimates can only be interpreted in the context of defined transcriptional units (for example, comparison of the expression levels of exons, specific transcripts, or genes). When RNA is being sequenced from a species for which a well-annotated genome is available, RNA sequencing reads can be aligned to the previously defined transcriptome and expression levels can be estimated based on the number of aligned reads. The problem is that only few genomes are well annotated.

When a genome is available but not well annotated, two approaches can be used to define transcriptional units. The first relies on the functional annotations from a closely related genome and this approach has to overcome the challenge of accurately defining orthology. A conservative definition of orthology, requiring high sequence similarity for assignments, risks excluding a large fraction of transcriptional units from the analysis, whereas relaxed criteria (i.e accepting weaker evidence for homology), can result in erroneous orthology assignments. The second approach is to align RNA sequencing reads to the genome sequence and *de novo* define expressed transcriptional units. This task is far from trivial, as it requires one to distinguish foreground expression levels from the background (such as sequencing reads corresponding to unspliced introns).

When a genome sequence is not available, *de novo* transcriptome assembly is required. This is a particularly challenging task, because it does not rely on an alignment of the sequencing reads to a known genome. Despite this technical challenge, for the purpose of comparing expression levels across species, the data obtained via *de novo* transcriptome assembly are expected to have the same properties as those obtained from defining transcriptional units based on aligning RNA sequence reads to a genome. Thus, transcriptome assembly is an attractive approach for studies on any species for which genome sequences are not yet available. That said, with the rapid decrease in sequencing costs and the corresponding increase in sequencing capacity, it might be reasonable to expect that sequencing large (e.g., mammalian) genomes may not be a prohibitive enterprise in the near future.

Box 2 – The signatures of natural selection on gene regulation

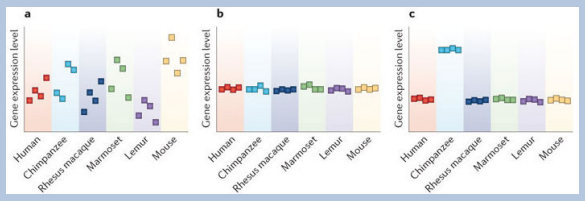
How can one distinguish between different modes of gene expression evolution? One approach is to look for departures from a null model of a given evolutionary scenario. At the sequence level, the most commonly used null is the neutral model, which proposes that some alleles are strongly deleterious, are subjected to strong purifying selection and thus are never seen in a sample, whereas the alleles that do segregate in the population are selectively neutral[109, 110]. In the case of a quantitative phenotype such as gene expression levels, evolutionary constraint is likely to take the form of stabilizing selection, which maintains a constant mean and reduces the variance of the trait[111, 112]. However, as discussed in the main text, it is difficult to specify the expectations under the null for non-model species. An alternative is to use an empirical approach to identify gene expression patterns that likely have evolved under natural selection.

For example, if gene regulation evolves under stabilizing selection genes are expected to show little variation in expression levels within and between species. In contrast, under directional selection in a particular lineage, genes are expected to show a significant shift in the mean expression level in that one lineage and show little variation in expression levels among individuals within a species[24]. This is illustrated schematically in the figure: gene expression levels (y-axis) are plotted for four individuals from each of six mammalian species. In panel A, variation in gene expression level is high both within and between species. This might not be unexpected given that it is difficult to stage tissues and to minimize environmental effects on gene regulation in a comparative study. In panel B, little variation in gene expression levels is observed both within and between species. The most likely explanation for such a pattern, especially in the face of the technical limitations associated with comparative studies using non-model organisms, is that gene regulation evolves under stabilizing selection. The pattern shown in panel C indicates a change in gene expression level in the chimpanzee lineage, which is consistent with directional selection on gene regulation in chimpanzee. However, alternative explanations - such as lineage-specific relaxation of evolutionary constraint, or lineage-specific difference in environment - are difficult to exclude.

The inference of selection relies on the ranking of expression level variation within and between species, not direct evidence for the presence or absence of natural selection. Although statistical analyses are typically used to rank genes based on their gene expression patterns, this ranking-based approach should be considered heuristic and model-free. It is difficult to apply less heuristic approaches to the comparative analysis of gene expression levels in primates because one cannot directly study the mutational input for gene expression variation in these species, nor is one able to experimentally establish levels of gene expression divergence that indicate the action of natural selection rather than low mutational input.

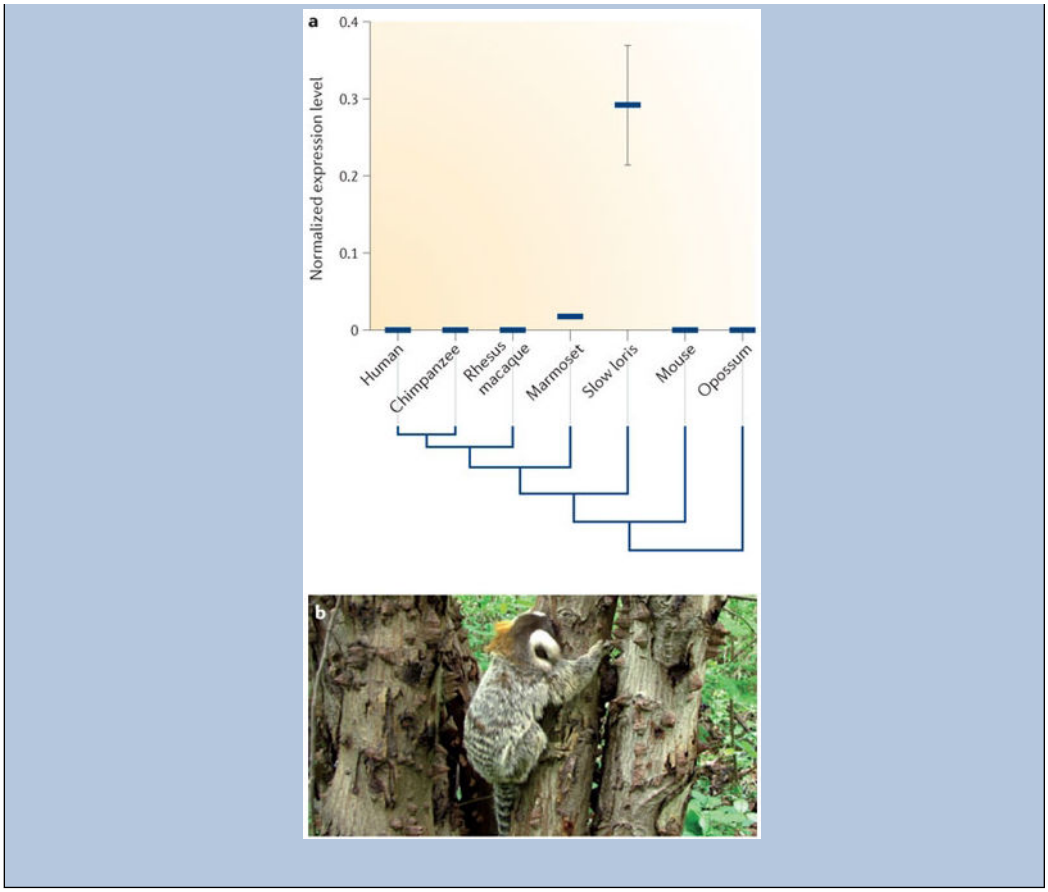
Similar empirical approaches are used in other types of genome-wide data analyses, for example, in scanning sequence data for evidence of recent natural selection on specific genes[113-116]. The general rationale is that genomic regions or genes ranked at the top of the list have nucleotide diversity or expression patterns that provide the most compelling evidence for the action of natural selection. It is therefore expected that genes at the top

of the list would be enriched for true targets of recent natural selection. It is recognized, however, that not all genomic regions at the top of list (regardless of the cutoff chosen) are indeed targets of natural selection, and conversely, not all true targets of natural selection will be at the top of the list[117, 118].



Box 3 - Inter-species regulatory differences and ecological adaptation: a case study

Comparative studies of gene expression levels might reveal the molecular signatures of ecological adaptations. An illustrative example of this is provided by the work of Perry and colleagues[16] who found that the expression levels of short chain dehydrogenase/ reductase family 16C, member 5 (*SDR16C5*) (panel A; y-axis) are elevated in the livers of marmoset and slow loris compared to all other studied primates (in livers from most other primates, the expression of this genes could not be detected). SDR16C5, an epidermal retinol dehydrogenase, is involved in the first, rate-limiting step of retinol (Vitamin A) metabolism. Retinol is a derivative of isoprene, the monomer of latex. Slow lorises and Marmosets feed extensively on tree exudates[119, 120], which may include gums, saps, and latex; a marmoset gouging tree bark is shown in panel B. Among the species considered in this study[16], only marmosets and slow lorises have apparent craniofacial adaptations for tree gouging. It is not known how exudates are digested in primates, but this process is thought to be aided by bacterial fermentation in the gut. In this case, there may be large quantities of the digestive products, such as retinol, absorbed through the large intestine, which may then be filtered by the liver. The intermediate-to-high expression levels of *SDR16C5* exclusively in the liver tissues of slow loris and marmoset could represent convergent adaptation against the fitness-reducing effects of vitamin A toxicity. Of course, such hypotheses based on single-gene observations should be considered highly tenuous. Nevertheless, this information may be valuable if it ultimately leads to further study and a better understanding of diet-related adaptations and evolutionary ecology in primates. The image is kindly provided by Ana Karinne Lima, Data for panel A is from reference 16.



Online 'at-a-glance' summary

- The hypothesis that differences in gene regulation play an important role in speciation and adaptation is more than 40 years old.
- RNA sequencing (RNA-seq) allows one to measure and compare gene expression levels across practically any combination of species at unprecedented resolution.
- Comparative studies of gene expression levels in all species studied to date provide compelling evidence that most gene regulatory patterns evolve under evolutionary constraint.
- It is more difficult to infer the action of positive (directional) selection on gene regulation than the action of stabilizing selection, especially in non-model species such as humans and non-human apes where environmental and genetic effects might be confounded.
- Inter-species differences in epigenetic markers can likely explain a substantial fraction of gene expression differences between species.
- Because a broad range of experimental manipulations are possible in model organisms, studies that focus on model species can move beyond simple comparisons of gene expression and offer deep insights into the causal relationship between regulatory changes and phenotypic evolution.
- Functional studies in model systems can often shed light on the adaptive phenotypes that were affected by regulatory changes between humans and other primates. Some phenotypes, though (e.g., the development of language) are inherently difficult to study using model species.
- One might be able to use iPSCs derived differentiated cells from humans and non-human primates to functionally test for the outcomes of inter-species differences in gene regulation.

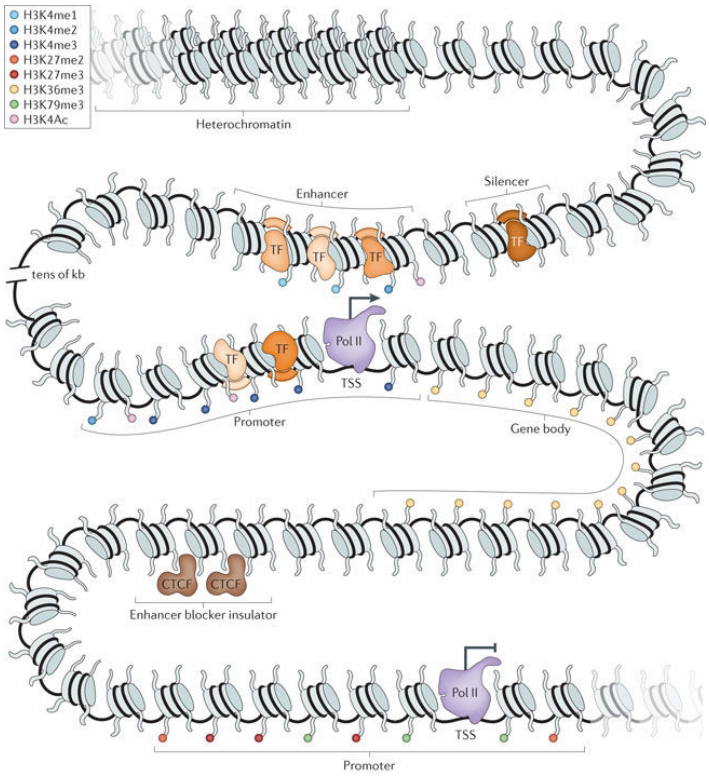


Figure 1. Regulatory mechanisms that can be investigated using comparative genomic approaches

Changes in a large number of genetic and epigenetic regulatory mechanisms can underlie inter-species differences in gene expression levels. Second-generation sequencing technologies allow us to obtain genome-wide profiles of transcription factor binding and epigenetic markers and thus identify correlations between variation in gene expression and variation in regulatory mechanisms. Using this paradigm, current studies are actively estimating the relative contribution of changes in different mechanisms to regulatory evolution, including chromatin accessibility (using DNaseI sequencing), Nucleosome positions (using MNase sequencing), transcription factor binding (using Chip-seq), promoter methylation profiling (using microarrays or bisulfite sequencing), and a number of histone modification profiles (using ChIP-seq). Figure is modified, with permission, from reference 121 **[Copy-ed: permission received.]**

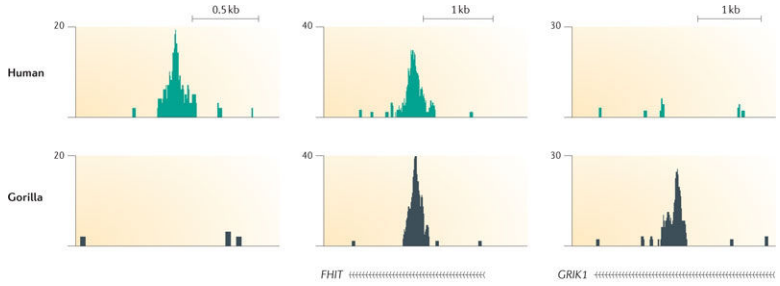


Figure 2. Inter-species differences in transcription factor binding

Ward, Odom, and colleagues have performed and analyzed comparative ChIP-seq experiments for the transcriptional regulator CTCF in human and gorilla cell 122 After ChIP-seq reads are mapped to the respective genomes, the resulting peaks (read counts are plotted on the y-axis) indicate the locations of chromatin enrichment and hence of CTCF binding. Examples are shown of a site bound in humans but not in gorilla within 2 kb of the *GPR88* gene (G-protein coupled receptor expressed in striatum) (this gene is not shown on the figure), a shared site at *FHIT* (triphosphate hydrolase possible tumour suppressor) and a site bound in gorillas but not in humans at *GRIK1* (glutamate receptor subunit involved in neurotransmission). The data for this figure are from reference 122.