Published in final edited form as:
Curr Opin Genet Dev. 2014 December ; 0: 68–74. doi:10.1016/j.gde.2014.08.010.

Comparative studies of gene regulatory mechanisms

Athma A. Pai[1] and **Yoav Gilad**[2]

[1]Department of Biology, Massachusetts Institute of Technology

[2]Department of Human Genetics, University of Chicago

Abstract

It has become increasingly clear that changes in gene regulation have played an important role in adaptive evolution both between and within species. Over the past five years, comparative studies have moved beyond simple characterizations of differences in gene expression levels within and between species to studying variation in regulatory mechanisms. We still know relatively little about the precise chain of events that lead to most regulatory adaptations, but we have taken significant steps towards understanding the relative importance of changes in different mechanisms of gene regulatory evolution. In this review, we first discuss insights from comparative studies in model organisms, where the available experimental toolkit is extensive. We then focus on a few recent comparative studies in primates, where the limited feasibility of experimental manipulation dictates the approaches that can be used to study gene regulatory evolution.

Introduction

The controversy over whether changes in gene regulation are disproportionally important in speciation and adaptation relative to changes in protein coding sequences has not yet been resolved [1–3]. Regardless, it has become clear that across a wide range of species, a large number of adaptations can be explained by changes in gene expression levels [4–9]. Similarly, the related question of whether most inter-species differences in gene expression levels have evolved neutrally or were subjected to selective pressures is still unanswered [10]. However, comparative and functional studies of gene expression levels have resulted in a better appreciation of the patterns of regulatory variation within and between species [11–13], and it is now possible to point to subsets of genes whose expression have likely evolved under lineage-specific directional selection [11,12]. The next natural step is to focus on characterizing the underlying regulatory mechanisms.

Broadly speaking, differences in gene expression levels are due to changes in *cis* and/or *trans* regulatory mechanisms [14]. Regulatory elements that act in *cis* (namely, elements that

Corresponding author: YG (gilad@uchicago.edu).

influence allele-specific regulation) include binding sites for transcription factors and small RNAs, as well as sites for chromatin modifiers and marks that determine nucleosome positioning or the degree of chromatin accessibility. Regulatory elements that act in *trans* (namely, affect the regulation of both alleles) include transcription factors and small RNAs, as well as enzymes that modify chromatin and establish epigenetic marks.

Both *cis* and *trans* elements can regulate steady-state gene expression levels by affecting the rates of either transcription or RNA decay. Yet, it has been shown that variation in transcription rates likely accounts for the majority of the overall variation in steady-state transcript levels [15]. Moreover, it has been argued that changes in *cis* might underlie phenotypic adaptations more often than changes in *trans*, since changes to *cis* regulatory elements could be restricted to specific spatial and temporal consequences while changes in *trans* are likely to be associated with general pleiotropic, and often deleterious, effects [14].

Consistent with this notion, we know of a few dozen cases of adaptations in different species that could be explained by changes in gene expression due to genetic variation in *cis* regulatory elements (e.g. [4–9]). Yet we know of a very small number of cases of species-specific regulatory adaptations that can be explained by changes in *trans* elements [16]. In humans, for example, one of the best-characterized cases of possible regulatory adaptation through a *cis* element involves the human-accelerated non-coding sequence 1 (HACNS1), an enhancer region in which human-specific fixed substitutions were shown to drive limb bud expression of nearby genes with possible consequences for human limb development [17]. In contrast, there are no convincing reports yet of human-specific *trans* regulatory adaptations (though one could arguably consider the accelerated evolution of the human *FOXP2* gene as a possible example [18]). This discrepancy might be partly explained by the inherent difficulty of studying the consequences of suspected adaptive changes in *trans* elements.

Comparative studies of cis and trans elements

In model organisms and species in which experimentation is feasible, the focus of comparative studies is typically to uncover the genetic and gene regulatory basis for phenotypic adaptations (these studies are not reviewed here). However, a few studies took advantage of the ability to design specific experiments in model organisms to directly address the question of the relative importance of changes in *cis* and *trans* regulatory mechanisms to the evolution of gene expression. The commonly used approach is to compare RNA sequencing based estimates of allele-specific expression (ASE) levels in F1 hybrids to overall gene expression levels in the homozygote F0 parents.

Using this study design (Figure 1), it is possible to infer whether gene expression differences between the parents are due to *cis*-acting genetic differences that affect allele-specific expression or due to genetic differences that affect both alleles in the F1s, namely differences in *trans* elements. Though the approach does not allow one to easily identify the specific causal regulatory sequence elements, studies using this paradigm took some of the first steps towards deciphering the logic of gene regulatory evolution. Such studies in both flies and mice have suggested that most gene expression differences between strains or closely related species are due to changes in *cis* regulatory elements [19,20]. These studies

have also uncovered substantial and previously under-appreciated contributions from changes in *trans* elements (often acting in combination with changes in *cis*) to inter-strain and inter-species regulatory variation [21–23].

In contrast to model organisms, comparative studies in humans and other primates have used indirect approaches to study the relative importance of changes in different regulatory mechanisms, since direct experimentation or hybrid approaches are impossible for ethical and practical reasons. Indeed, changes in *cis* and *trans* regulatory elements in primates were generally inferred based on comparative chromatin immunoprecipitation followed by high-throughput sequencing (ChIP-seq) data. A few recent studies used ChIP-seq data to compare the binding profiles of individual transcription factors (TFs) to identify inter-species differences in binding, and binding site turnover, for specific factors [24–27]. These studies have generally revealed little conservation of TF binding profiles across even closely related primate species [24–26,28]. Turnover of TF binding sites is frequent, and even when a TF is bound to orthologous promoter regions across species, it is often not bound to the exactly orthologous regulatory locus [24–26,28].

A potential caveat is that nearly all of the comparative ChIP-seq studies published to date have focused on TFs with broad functions, whose binding patterns are extensive (tens of thousands of binding sites genome wide; [25–27]). It is likely that most of these binding events are not directly functional (in the sense that they do not lead to changes in gene regulation [29]) and hence are not expected to be conserved.

An alternative approach is to use comparative profiling of chromatin accessibility, which measures broader differences in putative regulatory elements. This can be done using techniques such as the DNaseI hypersensitivity assay (DNase-seq), which is a genome-wide chromatin accessibility assay that involves the digestion of DNA in regions of open chromatin by the DNaseI enzyme followed by high-throughput sequencing of resulting fragments of accessible regions. DNaseI cleavage sites in open chromatin mark regions that are likely to be regulatory active [30]. Moreover, chromatin accessibility can be used to simultaneously infer the binding of many transcription factors, using DNaseI footprints within accessible regions, using a single assay per individual [31]. The extent of chromatin accessibility as assayed by DNase-seq is generally correlated with gene expression differences across genes and individuals within a species [32–34]. In a comparative context, Shibata *et al.* showed that DNaseI sensitivity differences might explain a modest proportion of differential expression across primates [35].

It should be noted that incomplete power to detect binding events, in either the comparative ChIP-seq or DNase-seq studies, can result in an inflation of apparent inter-species differences (or lack of conservation).

Explaining inter-species variation in gene expression levels

With the advent of better and cheaper high-throughput sequencing technologies, it became possible to characterize genome-wide variation within and between species in a large number of genetic and epigenetic regulatory mechanisms. Although the ultimate goal remains to be able to ‘read the code’, namely to figure out how the sequences at regulatory

elements determine gene expression patterns, an intermediate aim of studies of regulatory evolution is to identify the changes in specific regulatory mechanisms that explain inter-species variation in gene expression levels.

Unfortunately, comparative genome-wide studies are limited in their ability to infer direct causality because they rely on correlations between datasets. Indeed, the general approach has been to characterize inter-species differences in gene expression levels using RNA sequencing, along with variation in one or more regulatory mechanisms. The assumption, based on the central dogma, is that gene expression levels are the output, and changes in regulatory interactions and/or mechanisms are the cause for differences in the output levels. Correlations between inter-species variation in gene expression levels and differences in regulatory mechanisms between species are therefore interpreted as likely to indicate causality. Almost all comparative studies in primates to date, which used this general approach to 'explain' inter-species differences in gene expression levels, have focused on mechanisms that regulate or affect transcription.

For instance, comparative studies in primates that assessed the regulatory impact of duplications in the genome – a mechanism that can indirectly affect overall gene expression levels - have found that genes within species-specific segmental duplications or gains in copy number tend to be expressed at a higher level than in the other species [36,37]. It was estimated that 2–8% of differentially expressed genes between humans and chimpanzees might be due to inter-species differences in segmental duplications [38].

A similar approach was used to assess the degree of inter-species gene expression differences that are associated with changes in epigenetic modifications between species. For instance, a study that focused on histone marks collected gene expression data and H3K4me3 histone modification profiles (a marker of active promoters) in lymphoblastoid cell lines (LCLs) from humans, chimpanzees, and rhesus macaques [39]. They found strong evidence for conservation of H3K4me3 localization in primates. Highly expressed genes were more likely than genes expressed at low levels to have the histone modification near their transcription start site (TSS). Correspondingly, there was an enrichment of inter-species differences in H3K4me3 marks at the TSS of genes that are differentially expressed between species; an estimated 7% of gene expression differences between the LCLs of humans, chimpanzees, and rhesus macaques were associated with changes in the status of H3K4me3 histone modifications [39].

In turn, a comparative methylation study examined DNA methylation patterns in livers, hearts, and kidneys from multiple humans and chimpanzees, using tissue samples for which genome-wide gene expression data were also available [40]. They found that tissue-specific methylation patterns are often conserved between humans and chimpanzees. Inter-species differences in gene expression levels were often associated with corresponding differences in methylation levels, and accounting for the levels of DNA methylation often decreased evidence for differential expression between the species (Figure 2). This study estimated that in the tissues considered, inter-species differences in promoter methylation might underlie as much as 12%–18% of differences in gene expression levels between humans and chimpanzees [40].

Considering additional mechanisms

A subset of inter-species differences in gene expression levels may be explained by changes in transposable elements. Ward *et al.* studied the transcriptional potential and epigenetic regulation of silenced repetitive regions in a transchromosomic mouse strain containing an almost complete copy of the human chromosome 21 [41]. Previously, using this transchromosomic mouse model, the same group showed that transcriptional control of genes on human chromosome 21 was largely driven by *cis*-acting sequence rather than the mouse *trans*-environment [42]. In this latest study, Ward and colleagues showed that within the mouse *trans*-environment, many primate and human-specific transposable elements were differentially marked by both DNA methylation and activating histone modifications and had the latent potential to regulate the expression levels of nearby genes [41].

Another subset of gene expression differences between species may be due to changes in the regulation of alternative splicing. Genome-wide studies of splicing patterns across multiple tissues from many mammalian species showed that species-specific splicing effects can be even more pronounced than tissue-specific effects (in contrast to overall gene expression profiles, which exhibit stronger tissue-specific signatures [43,44]). Furthermore, mammalian splicing patterns are often regulated by *cis* changes in motifs bound by RNA binding proteins with seemingly conserved regulatory functions [43]. These results implicate alternative splicing and other un-characterized post-transcriptional mechanisms as rich sources of novel regulatory variation.

Put together, the collection of comparative studies of gene regulatory mechanisms builds up to a highly complex, yet far from complete, description of the relative impact of various mechanisms on the evolution of gene expression. As we struggle to comprehend the vast amounts of comparative data pertaining to regulatory mechanisms that affect steady-state gene expression, new studies suggest that much of the variation in transcript level may not actually be translated into differences at the protein level. For example, a recent study comparing transcript and protein expression levels in primates found evidence for buffering of mRNA differences at the protein level [45]. These observations point to an even more complex relationship between changes in gene regulation and phenotypic adaptation [21,45].

Summary

The few example studies we discussed represent a growing body of comparative genomic work aimed to understand the mechanistic basis for inter-species differences in gene expression levels. We note, however, that practically all of these studies consider the contribution of individual regulatory mechanisms. Many regulatory interactions are highly correlated with each other and the causal order of events among such correlated interactions is still unknown. This poses two general challenges for future comparative studies. First, while the central dogma provides a strong foundation for the inference that (in most cases) gene expression is the output caused by changes in regulatory mechanisms, no such foundation exists with which to infer the single causative change that starts the regulatory cascade. Second, while it is possible to estimate the marginal contribution of changes in different regulatory mechanisms to differences in gene expression levels, it is unclear how to model higher order interactions between regulatory relationships.

In our opinion, it is unlikely that additional comparative genomic studies with similar study designs can provide answers to these questions. Instead, functional studies in cell lines or species in which experimentation is feasible are needed to directly study causality and resolve the mechanistic relationships between regulatory interactions.

Functional studies in systems where manipulation and experimentation is feasible are also needed to connect the regulatory changes to inter-species differences in complex phenotypes. As we discussed in the introduction, the functional consequences of inter-primate differences in gene regulation have been studied using genetic manipulation in model species (mainly mice), but this approach is laborious: Only a handful of worked-out examples are known to date. Moreover, even at its best, these approaches require a 'leap of faith' (because direct experimentation is not possible) when carrying over the observations in mice to the biology of primates. As a result, at the moment, we do not yet know the precise functional consequences of any gene regulatory changes between primate species, even though many identified patterns are consistent with the action of natural (stabilizing or directional) selection and hence are likely to be of functional importance.

Comparative studies in cell lines, for example in induced pluripotent stem cells and their differentiated cell types, can potentially provide better insights. While it would still be impossible to examine the regulatory basis of inter-primate differences in some of the most fascinating human-specific traits (such as the development of language or other cognitive traits), one could study a wide range of important cellular phenotypes. For instance, comparative studies of metabolic differences, response to stress, susceptibility to disease, and variation in drug toxicity can all be performed using cell line systems. The genetic basis for and regulatory pathways underlying inter-species differences in such traits can be directly manipulated and causality can be inferred.

Acknowledgments

This work was supported by NIH grant MH084703 to YG and Jonathan Pritchard, and a postdoctoral fellowship from the Jane Coffin Childs Foundation to AAP.

References

1. King M-C, Wilson A. Evolution at Two Levels in Humans and Chimpanzees. Science. 1975; 188:107–116. [PubMed: 1090005]
2. Hoekstra HE, Coyne JA. THE LOCUS OF EVOLUTION: EVO DEVO AND THE GENETICS OF ADAPTATION. Evolution. 2007; 61:995–1016. [PubMed: 17492956]
3. Wray GA. The evolutionary significance of cis-regulatory mutations [Internet]. Nat Rev Genet. 2007; 8:206–216. [PubMed: 17304246]
4. Shapiro MD, Marks ME, Peichel CL, Blackman BK, Nereng KS, Jonsson B, Schluter D, Kingsley DM. Genetic and developmental basis of evolutionary pelvic reduction in threespine sticklebacks. Nature. 2004; 428:717–723. [PubMed: 15085123]
5. Abzhanov A, Protas M, Grant BR, Grant PR, Tabin CJ. Bmp4 and Morphological Variation of Beaks in Darwin's Finches. Science. 2004; 305(5689):1462–1465. [PubMed: 15353802]
6. Jeong S, Rebeiz M, Andolfatto P, Werner T, True J, Carroll SB. The evolution of gene regulation underlies a morphological difference between two Drosophila sister species. Cell. 2008; 132:783–793. [PubMed: 18329365]

7. Gompel N, Prud'homme B, Wittkopp PJ, Kassner VA, Carroll SB. Chance caught on the wing: cis-regulatory evolution and the origin of pigment patterns in Drosophila. Nature. 2005; 433:418–487.

8. Chan YF, Marks ME, Jones FC, Villarreal G, Shapiro MD, Brady SD, Southwick AM, Absher DM, Grimwood J, Schmutz J, et al. Adaptive Evolution of Pelvic Reduction in Sticklebacks by Recurrent Deletion of a Pitx1 Enhancer. Science. 2010; 327:302–305. [PubMed: 20007865]

9. Werner T, Koshikawa S, Williams TM, Carroll SB. Generation of a novel wing colour pattern by the Wingless morphogen. Nature. 2010; 464:1143–1148. [PubMed: 20376004]

10. Romero IG, Ruvinsky I, Gilad Y. Comparative studies of gene expression and the evolution of gene regulation. Nat Rev Genet. 2012; 13:505–516. [PubMed: 22705669]

11. Khaitovich P, Hellmann I, Enard W, Nowick K, Leinweber M, Franz H, Weiss G, Lachmann M, Pääbo S. Parallel patterns of evolution in the genomes and transcriptomes of humans and chimpanzees. Science. 2005; 309:1850–1854. [PubMed: 16141373]

12. Gilad Y, Oshlack A, Smyth GK, Speed TP, White KP. Expression profiling in primates reveals a rapid evolution of human transcription factors. Nature. 2006; 440:242–245. [PubMed: 16525476]

13. Schadt EE, Monks SA, Drake TA, Lusis AJ, Che N, Colinayo V, Ruff TG, Milligan SB, Lamb JR, Cavet G, et al. Genetics of gene expression surveyed in maize, mouse and man. Nature. 2003; 422:297–302. [PubMed: 12646919]

14. Prud'homme B, Gompel N, Carroll SB. Emerging principles of regulatory evolution. Proc Natl Acad Sci USA. 2007; 104 (Suppl 1):8605–8612. [PubMed: 17494759]

15. Pai AA, Cain CE, Mizrahi-Man O, De Leon S, Lewellen N, Veyrieras J-B, Degner JF, Gaffney DJ, Pickrell JK, Stephens M, et al. The contribution of RNA decay quantitative trait loci to inter-individual variation in steady-state gene expression levels. PLoS Genet. 2012; 8:e1003000. [PubMed: 23071454]

16. Baker CR, Booth LN, Sorrells TR, Johnson AD. Protein Modularity, Cooperative Binding, and Hybrid Regulatory States Underlie Transcriptional Network Diversification. Cell. 2012; 151:80–95. [PubMed: 23021217]

*17. Prabhakar S, Visel A, Akiyama JA, Shoukry M, Lewis KD, Holt A, Plajzer-Frick I, Morrison H, Fitzpatrick DR, Afzal V, et al. Human-specific gain of function in a developmental enhancer. Science. 2008; 321:1346–1350. One of very few worked-put examples of some functional consequences of a change in a regulatory element along the human lineage. [PubMed: 18772437]

18. Enard W, Przeworski M, Fisher SE, Lai CSL, Wiebe V, Kitano T, Monaco AP, Paabo S. Molecular evolution of FOXP2, a gene involved in speech and language. Nature. 2002; 418:869–872. [PubMed: 12192408]

*19. Wittkopp PJ, Haerum BK, Clark AG. Evolutionary changes in cis and trans gene regulation. Nature. 2004; 430:85–88. An intelligent study design allowed the authors to estimate the relative importance of changes in cis and trans regulatory elements to the evolution of gene expression in flies. [PubMed: 15229602]

20. Wittkopp PJ, Haerum BK, Clark AG. Regulatory changes underlying expression differences within and between Drosophila species. Nat Genet. 2008; 40:346–350. [PubMed: 18278046]

21. McManus CJ, Coolon JD, Duff MO, Eipper-Mains J, Graveley BR, Wittkopp PJ. Regulatory divergence in Drosophila revealed by mRNA-seq. Genome Research. 2010; 20:816–825. [PubMed: 20354124]

22. Coolon JD, McManus CJ, Stevenson KR, Graveley BR, Wittkopp PJ. Tempo and mode of regulatory evolution in Drosophila. Genome Research. 2014; 24:797–808. [PubMed: 24567308]

23. Gonçalves Â, Leigh-Brown S, Thybert D, Stefflova K, Turro E, Flicek P, Brazma A, Odom DT, Marioni JC. Extensive compensatory cis-trans regulation in the evolution of mouse gene expression. Genome Research. 2012; 22:2376–2384. [PubMed: 22919075]

24. Schmidt D, Wilson MD, Ballester B, Schwalie PC, Brown GD, Marshall A, Kutter C, Watt S, Martinez-Jimenez CP, Mackay S. Five-vertebrate ChIP-seq reveals the evolutionary dynamics of transcription factor binding. Science. 2010; 328:1036–1040. [PubMed: 20378774]

25. Schmidt D, Schwalie PC, Wilson MD, Ballester B, Gonçalves Â, Kutter C, Brown GD, Marshall A, Flicek P, Odom DT. Waves of Retrotransposon Expansion Remodel Genome Organization and CTCF Binding in Multiple Mammalian Lineages. Cell. 2012; 148:335–348. [PubMed: 22244452]

26. Kutter C, Brown GD, Gonçalves Â, Wilson MD, Watt S, Brazma A, White RJ, Odom DT. Pol III binding in six mammals shows conservation among amino acid isotypes despite divergence among tRNA genes. Nat Genet. 2011; 43:948–955. [PubMed: 21873999]
27. Scally A, Dutheil JY, Hillier LW, Jordan GE, Goodhead I, Herrero J, Hobolth A, Lappalainen T, Mailund T, Marques-Bonet T, et al. Insights into hominid evolution from the gorilla genome sequence. Nature. 2012; 483:169–175. [PubMed: 22398555]
28. Stefflova K, Thybert D, Wilson MD, Streeter I, Aleksic J, Karagianni P, Brazma A, Adams DJ, Talianidis I, Marioni JC, et al. Cooperativity and rapid evolution of cobound transcription factors in closely related mammals. Cell. 2013; 154:530–540. [PubMed: 23911320]
29. Cusanovich DA, Pavlovic B, Pritchard JK, Gilad Y. The functional consequences of variation in transcription factor binding. PLoS Genet. 2014; 10:e1004226. [PubMed: 24603674]
30. Boyle AP, Davis S, Shulha HP, Meltzer P, Margulies EH, Weng Z, Furey TS, Crawford GE. High-resolution mapping and characterization of open chromatin across the genome. Cell. 2008; 132:311–322. [PubMed: 18243105]
31. Pique-Regi R, Degner JF, Pai AA, Gaffney DJ, Gilad Y, Pritchard JK. Accurate inference of transcription factor binding from DNA sequence and chromatin accessibility data. Genome Research. 2011; 21:447–455. [PubMed: 21106904]
32. Xi H, Shulha HP, Lin JM, Vales TR, Fu Y, Bodine DM, McKay RDG, Chenoweth JG, Tesar PJ, Furey TS, et al. Identification and characterization of cell type-specific and ubiquitous chromatin regulatory structures in the human genome. PLoS Genet. 2007; 3:e136. [PubMed: 17708682]
33. Boyle AP, Song L, Lee B-K, London D, Keefe D, Birney E, Iyer VR, Crawford GE, Furey TS. High-resolution genome-wide in vivo footprinting of diverse transcription factors in human cells. Genome Research. 2011; 21:456–464. [PubMed: 21106903]
34. Degner JF, Pai AA, Pique-Regi R, Veyrieras J-B, Gaffney DJ, Pickrell JK, De Leon S, Michelini K, Lewellen N, Crawford GE, et al. DNase I sensitivity QTLs are a major determinant of human expression variation. Nature. 2012; 482:390–394. [PubMed: 22307276]
35. Shibata Y, Sheffield NC, Fedrigo O, Babbitt CC, Wortham M, Tewari AK, London D, Song L, Lee B-K, Iyer VR, et al. Extensive Evolutionary Changes in Regulatory Element Activity during Human Origins Are Associated with Altered Gene Expression and Positive Selection. PLoS Genet. 2012; 8:e1002789. [PubMed: 22761590]
36. Khaitovich P, Muetzel B, She X, Lachmann M, Hellmann I, Dietzsch J, Steigele S, Do H-H, Weiss G, Enard W, et al. Regional patterns of gene expression in human and chimpanzee brains. Genome Research. 2004; 14:1462–1473. [PubMed: 15289471]
37. Iskow RC, Gokcumen O, Abyzov A, Malukiewicz J, Zhu Q, Sukumar AT, Pai AA, Mills RE, Habegger L, Cusanovich DA, et al. Regulatory element copy number differences shape primate expression profiles. Proc Natl Acad Sci USA. 2012; 109:12656–12661. [PubMed: 22797897]
38. Blekhman R, Oshlack A, Gilad Y. Segmental duplications contribute to gene expression differences between humans and chimpanzees. Genetics. 2009; 182:627–630. [PubMed: 19332884]
39. Cain CE, Blekhman R, Marioni JC, Gilad Y. Gene expression differences among primates are associated with changes in a histone epigenetic modification. Genetics. 2011; 187:1225–1234. [PubMed: 21321133]
40. 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 Genet. 2011; 7:e1001316. [PubMed: 21383968]
*41. Ward MC, Wilson MD, Barbosa-Morais NL, Schmidt D, Stark R, Pan Q, Schwalie PC, Menon S, Lukk M, Watt S, et al. Latent regulatory potential of human-specific repetitive elements. Molecular Cell. 2013; 49:262–272. Along with the initial description of the mouse line carrying a human chr21 (ref 42), this paper accentuated the power of direct experimentation to distinguish between the contributions of *cis*- or *trans*-acting elements to regulatory divergence. [PubMed: 23246434]
*42. Wilson MD, Barbosa-Morais NL, Schmidt D, Conboy CM, Vanes L, Tybulewicz VL, Fisher EM, Tavaré S, Odom DT. Species-specific transcription in mice carrying human chromosome 21. Science. 2008; 322:434–438. [PubMed: 18787134]

43. Barbosa-Morais NL, Irimia M, Pan Q, Xiong HY, Gueroussov S, Lee LJ, Slobodeniuc V, Kutter C, Watt S, Colak R, et al. The Evolutionary Landscape of Alternative Splicing in Vertebrate Species. Science. 2012; 338:1587–1593. [PubMed: 23258890]

44. Merkin J, Russell C, Chen P, Burge CB. Evolutionary Dynamics of Gene and Isoform Regulation in Mammalian Tissues. Science. 2012; 338:1593–1599. [PubMed: 23258891]

*45. Khan Z, Ford MJ, Cusanovich DA, Mitrano A, Pritchard JK, Gilad Y. Primate transcript and protein expression levels evolve under compensatory selection pressures. Science. 2013; 342:1100–1104. This study comparing gene expression and protein levels differences across primates highlighted our still meager understandings of the complex regulatory cascades governing the translation of mRNA into ultimate protein phenotypes. [PubMed: 24136357]

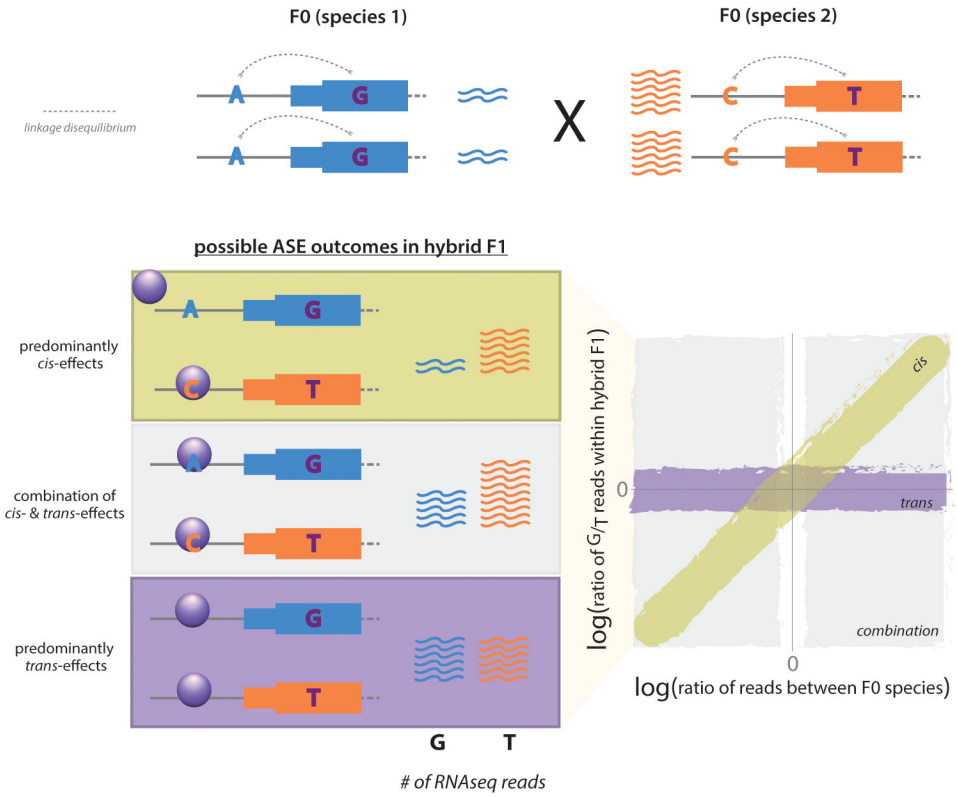

Figure 1. Studying allele-specific expression to understand the prevalence of *cis*- or *trans*-regulatory evolution

A common approach to characterizing the evolution of differential gene expression between species is to compare the differential expression between species to the differential allele-specific expression (ASE) of each F0 (parental) allele in an F1 hybrid organism, as pioneered by Wittkopp and colleagues in [19]. This figure illustrates the basic study design, which involves crossing two species with substantially diverged genomes such that a considerable amount of genes have fixed nucleotide differences in differentially expressed mRNA – represented by the blue G transcript and the orange T transcript from species 1 and species 2 respectively in the top panel. These polymorphisms might be in high linkage disequilibrium with often uncharacterized SNPs in regulatory regions. By comparing the ASE ratio of G to T alleles in the F1 organism to the differential expression between the F0 organisms, it is possible to assign each gene to one of three underlying regulatory modes (bottom panel). Predominantly *cis*-effects are represented by a constant ratio of G to T allelic expression within both the F0 and F1 organisms (yellow box), shown here as differential binding of a transcription factor to the upstream *cis*-regulatory polymorphism. Predominantly *trans*-effects are represented by equal expression of the G and T alleles in the F1 organism compared to the skewed ratio between the F0 organisms (purple box), shown here as binding of the *trans*-acting factor regardless of *cis*-regulatory sequence. Finally, genes regulated by a combination of *cis*- and *trans*- effects might show patterns of an overall altered (but still skewed) ratio between the G and T alleles in the F1 organism compared to the F0 organisms (grey box). These patterns are systematically assessed genome-wide by plotting the ratio of allele-specific expression in the hybrid F1 to the ratio of reads between

the two F0 species, as represented by the y- and x-axes respectively on the bottom right plot. Genes whose allele-specific expression patterns fall on the diagonal have often diverged due to predominantly *cis*-acting changes, while differentially expressed genes with no allele-specific expression in the F1 have often diverged due to predominantly *trans*-acting changes.

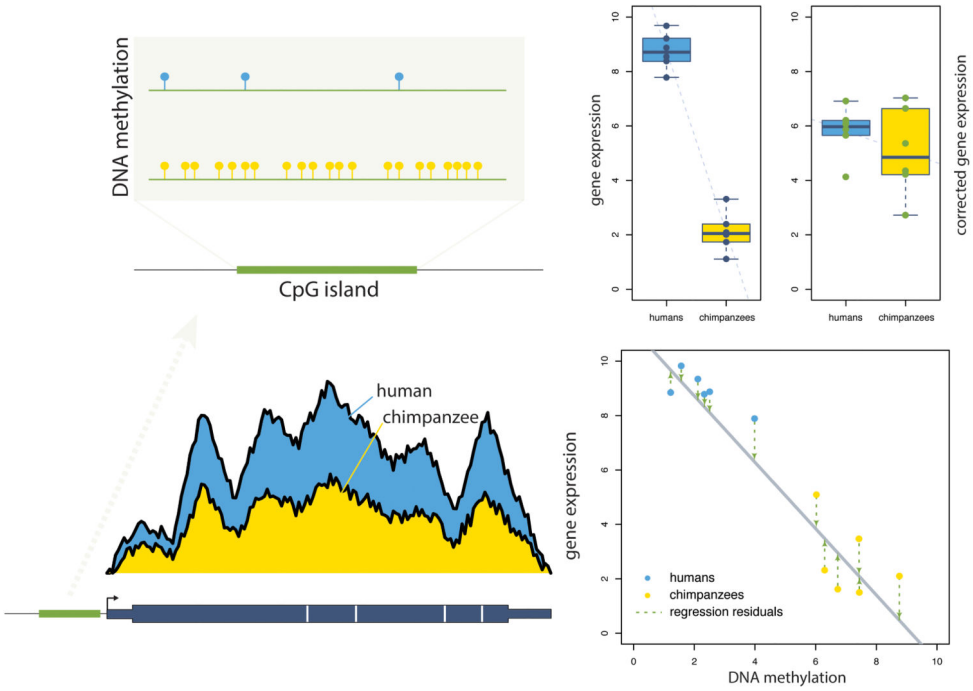

Figure 2. An approach for joint quantitative analysis of gene expression and regulatory data

A major goal in comparative studies of regulatory mechanisms is to understand the extent to which quantitative differences in regulatory marks underlie differential gene expression levels. One approach to computationally test this is to perform a joint analysis of the extent to which genes are differentially expressed after controlling for a regulatory mark. This example shows a gene that is both differentially methylated in an upstream CpG island (top left) and differentially expressed between human and chimpanzee individuals (bottom left and right panel, left boxplot – where each point represents one individual; human and chimpanzee data is represented in blue and yellow respectively). To quantitatively test the extent to which this correlation might underlie the differential expression, it is possible to either (1) perform a differential expression analysis on the residual gene expression levels after regressing out the DNA methylation levels (represented in the bottom right) or (2) include the DNA methylation levels as an explanatory variable in the differential expression analysis. In this example, regressing out methylation levels results in significantly reduced evidence for differential expression between human and chimpanzee (right panel, right boxplot), supporting the possibility that DNA methylation differences might be underlying the differential gene expression. Note that these approaches rely heavily on established directions of causality between gene regulatory processes from previously published literature.