Trends in Genetics

CellPress

Review

Evolution of Epigenetic Regulation in Vertebrate Genomes

Rebecca F. Lowdon,[1,*] Hyo Sik Jang,[1] and Ting Wang[1,*]

Empirical models of sequence evolution have spurred progress in the field of evolutionary genetics for decades. We are now realizing the importance and complexity of the eukaryotic epigenome. While epigenome analysis has been applied to genomes from single-cell eukaryotes to human, comparative analyses are still relatively few and computational algorithms to quantify epigenome evolution remain scarce. Accordingly, a quantitative model of epigenome evolution remains to be established. We review here the comparative epigenomics literature and synthesize its overarching themes. We also suggest one mechanism, transcription factor binding site (TFBS) turnover, which relates sequence evolution to epigenetic conservation or divergence. Lastly, we propose a framework for how the field can move forward to build a coherent quantitative model of epigenome evolution.

Trends

Epigenome evolution is characterized by variable conservation and divergence across the genome; within a clade (here vertebrates), rates of conservation or divergence are highly genome feature-specific.

TFBS turnover can mediate epigenome conservation or divergence.

Developmental genes are enriched in loci with divergent chromatin features, suggesting that rapid epigenome evolution may contribute novel regulatory mechanisms for lineage-specific characteristics.

Comparative Epigenomics as a Tool To Explore Epigenome Evolution

The epigenome is an integral part of genome biology, and comprises DNA modifications, most notably 5-methylcytosine (DNA methylation), histone post-translational modifications (PTMs), and nucleosome positioning (Figure 1). The epigenome is crucial for proper gene regulation [1], genome integrity [2], dosage compensation [3,4], and development [5] across eukaryotic phyla. Nevertheless, an empirical model of epigenome evolution has yet to be established. Decades of interrogating the chromatin remodeling of specific loci over development and across species provide early examples of **comparative epigenomics** (see Glossary), defined here as the comparison of epigenetic status between syntenic regions.

Comparative epigenomics is based on determining **epigenetic conservation**: two homologous sequences that host similar epigenetic modifications in homologous cell types (Figure 2). The homologous loci may be orthologous in distantly related species or paralogs in the same genome. It follows that epigenome comparison requires determination of sequence homology, epigenetic status, and biological homology between two species [6].

This review focuses on what comparative epigenomics has taught us about vertebrate epigenome evolution, although comparisons with invertebrate and plant epigenomes have been invaluable to build a full picture of epigenetic regulation [7–9]. In addition, the focus is confined to the use of comparative epigenetics, which can reveal epigenetic regulatory features by identifying regions of conserved and divergent epigenetic status across phyla, to understand gene regulation.

Lastly, the scope of this review is constrained by the scope of comparative epigenomics studies in existing literature. Figure 1 outlines common epigenetic marks and related assays that are

[1]Department of Genetics, Center for Genome Sciences and Systems Biology, Washington University School of Medicine, St. Louis, MO, USA

*Correspondence:
rebecca.lowdon@wustl.edu (R.F. Lowdon) and
twang@genetics.wustl.edu (T. Wang).

covered in this review, together with a representative example of the data and the interpretation of the epigenetic situation in a cartoon.

The arrival of high-throughput sequencing (HTS) technologies and genome-wide biochemistry experiments has moved the study of the epigenome into the 'omics' era. With HTS tools and databases of thousands of epigenome mapping experiments across thousands of eukaryotic individuals [10], the field can begin to create models of epigenome conservation and divergence and interpret the biological meaning behind these signals.

Epigenetic Evolution at Orthologs

Rooted in a strong theoretical foundation [11], comparative genomics enables the identification of conserved sequences, elucidating functional genomic elements [12,13]. However, not all functional genome regions are conserved [14,15], suggesting that other genomic features are responsible for adaptive gene regulation [16,17]. Two possible explanations for non-conserved functional elements are the limitation of sequence alignment algorithms [18] or that these non-conserved regions can serve as genuine species- or **lineage-specific** regulatory elements [16,19,20]. Accordingly, experimental approaches have shown that many non-conserved sequence elements are gene regulatory [21–23].

Pioneering work in comparative epigenetics detail the structure and function of chromatin and epigenetic modifications at orthologous loci across model organisms. Well-studied developmental loci, including the insulin-like growth factor 2 receptor locus, macrophage colony-stimulating factor, and the β-globin locus, exhibit conserved epigenetic status [24–26], transcription factor regulation [25–28], and function [25,28,29] across species (Box 1). Taken together, analysis of the sequence and epigenetic conservation at these loci suggests that epigenome comparison is a viable method for identifying elements modulating gene regulation.

From the above observations, it can be postulated that epigenetic features are correlated with underlying sequence features (Figure 2). This review presents evidence both for and against this hypothesis in an effort to establish a framework for epigenome evolutionary studies.

Relative DNA Methylation Conservation Across Sequence Contexts

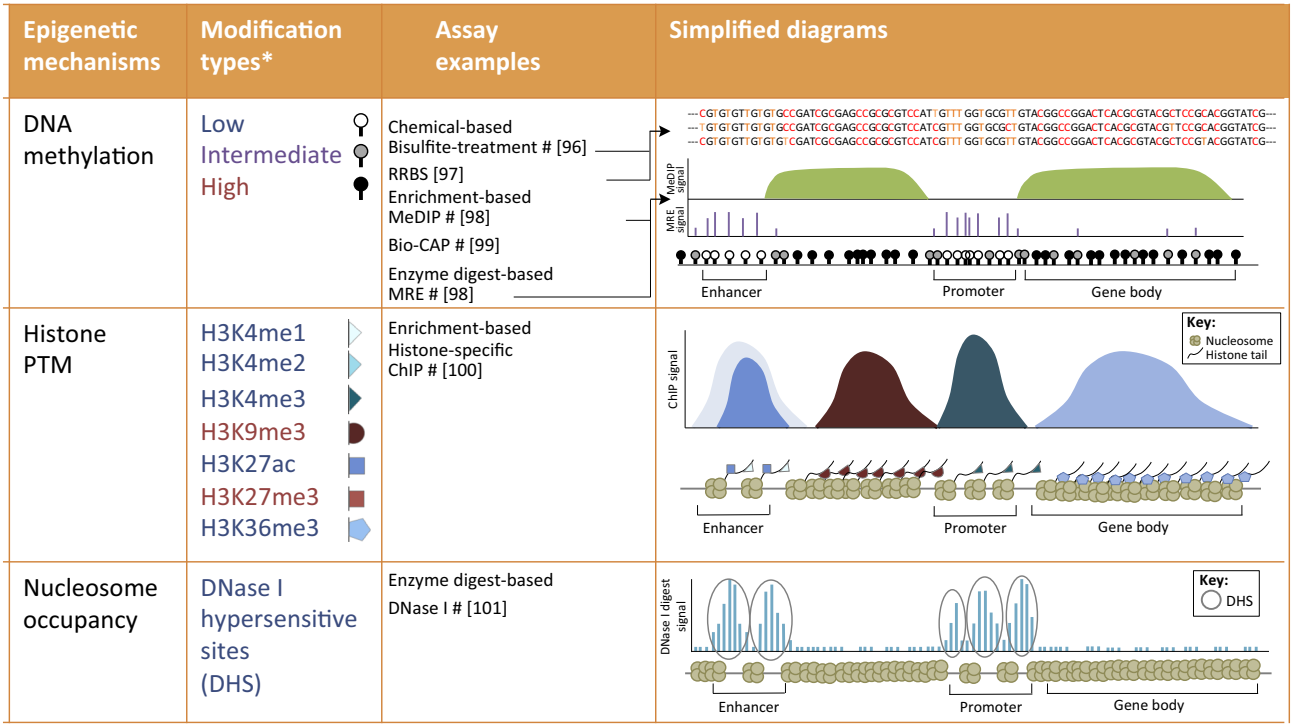
Analysis of epigenetic marks at paralogs allows epigenetic evolution to be studied without the confounding environmental variability that exists in inter-species comparisons [30]. In the human genome, 78% of paralogous CpGs had an absolute DNA methylation difference of 20% or less [30]. Thus, duplicons tend to retain their DNA methylation signature, supporting the hypothesis that epigenetic features are correlated with underlying sequence (Figure 2).

When comparing genome-wide DNA methylation levels between species, 70–74% and 80–82% similarities were found in peripheral blood and prefrontal cortex, respectively, in great ape somatic tissues [31,32]. Correlation coefficients from inter-species pairwise comparisons of whole-genome bisulfite sequencing (WGBS) data from primate blood samples show agreement with species phylogeny [32], suggesting that DNA methylation variation is related to sequence variation.

However, pairwise correlations of DNA methylation levels between species showed only moderate concordance at individual CpGs. For example, examination of primate peripheral blood samples using the Illumina Methylation450 array showed that 22% of probes covering orthologous CpGs were not significantly different among human, chimp, bonobo, gorilla, and orangutan (mean β value difference of <0.1) [31]. What accounts for individual CpG methylation level variance between species?

Glossary

Comparative epigenomics: comparison of epigenetic status between syntenic regions.
Epigenetic conservation: conservation of epigenetic marks at syntenic sequences; may or may not be coincident with conserved regulatory function of the locus.
Epigenetic divergence: different epi-marks at a syntenic loci; not necessarily indicative of differential function of the locus.
Lineage-specific: a genetic or epigenetic feature specific to an evolutionary lineage.
Transcription factor binding site (TFBS) turnover: nucleotide substitution that leads to a TFBS motif 'moving' along the DNA sequence in a locus, or 'transforming' into a different TFBS motif.



* = Red and blue terms correspond to repressive and active state, respectively.
# = These assays can be quantified by numerous techniques including, but not limited to, gel-imaging, targeted sequencing, RT-qPCR, microarrays and high-throughput sequencing.
[96–101] = Reference

Trends in Genetics

Figure 1. Dynamic Epigenetic Interactions. Innovations in sequence resolution and the identification of novel DNA-modifying mechanisms have provided novel opportunities to develop intricate techniques to explore epigenetic interactions. This review focuses on four unique, but usually complementary, epigenetic modifications that are universally shared across vertebrates. Different combinations of epigenetic modifications can define an active state, allowing expression of genes, or a repressive state, hindering gene expression, or an intermediate poised state that has potential to go in either direction. The biological function of individual epigenetic marks has been widely studied, but the combinatorial interactions across epigenetic modifications have still yet to be fully defined and understood. We diagram here simplified models to illustrate how results of epigenetics assays can be interpreted in the resolution of DNA context and chromatin context [96–101]. Key: *, red and blue terms correspond to repressive and active states, respectively; #, these assays can be quantified by numerous techniques including, but not limited to, gel imaging, targeted sequencing, RT-qPCR, microarrays, and high-throughput sequencing. Abbreviations: Bio-CAP, biotinylated CxxC affinity purification; MeDIP, methylated DNA immunoprecipitation; MRE, methylation-sensitive restriction enzyme; PTM, post-translational modification; RRBS, reduced representation bisulfite sequencing.

To test how DNA methylation status varies with sequence, regions of incomplete lineage sorting (ILS), where the sequence genealogy is different from the known species phylogeny, were used [32]. An example of an ILS would be an orthologous region in humans that is more similar in sequence to gorilla than chimp. The authors isolated 360 000 CpGs in ILS regions from human, chimp, gorilla, and orangutan. Strikingly, DNA methylation patterns over ILS regions followed the sequence relationships, suggesting a physical dependence of DNA methylation status at ILS regions on sequence variation [32]. In addition, most of the 570 regions of human-specific methylation were distal to transcription start-sites (TSSs) and showed accumulation of nucleotide substitutions [32], suggesting that methylome evolution may be coupled to sequence evolution at regulatory elements.

Relationships Between Histone PTM Conservation and Sequence Conservation

Comparative studies of histone PTMs show that not all orthologs have conserved epi-mark status. Comparative analysis of histone H3 dimethylation at lysine 4 (H3K4me2) using ChIP followed by microarray (ChIP-chip) over two orthologous loci in mouse and human lung fibroblasts showed that functional conservation of histone methylation did not correlate with

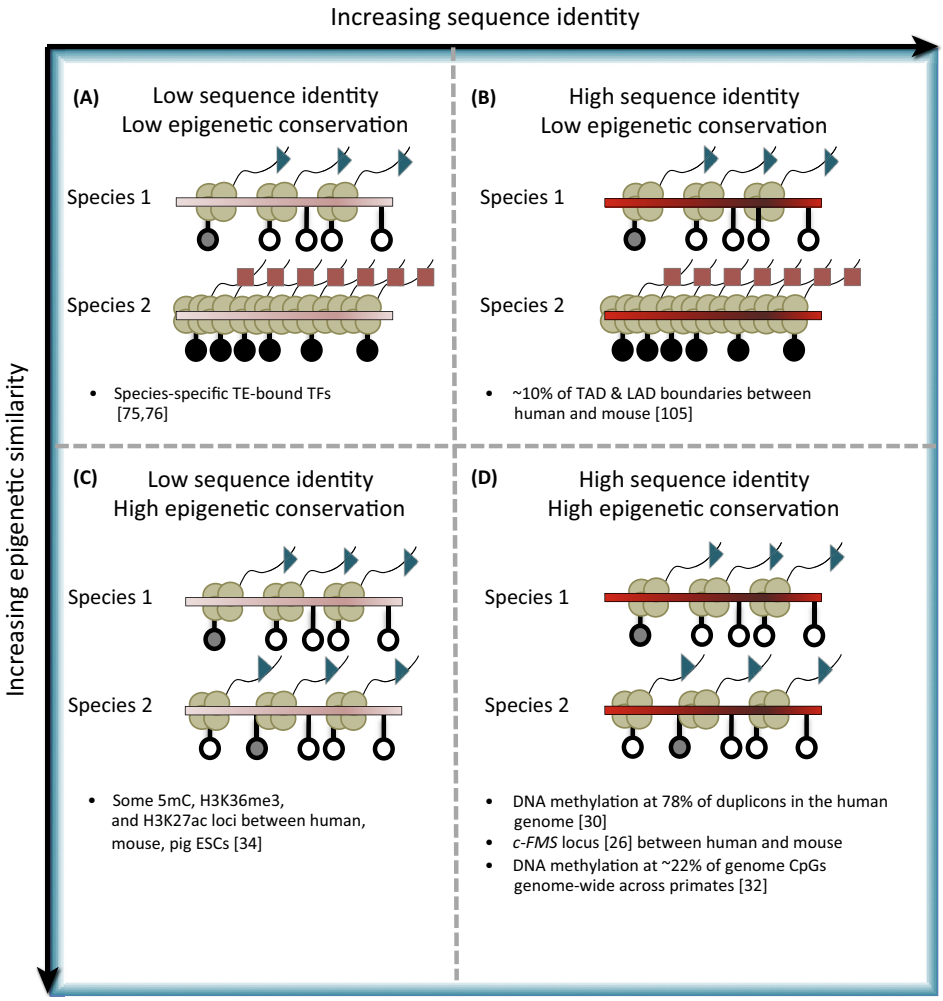


Figure 2. Genetic and Epigenetic Conservation Correlation. The degrees of sequence or epigenetic similarity between syntenic loci lie on a continuum. We define here epigenetic similarity as where the same epigenetic signal is present at orthologous loci in two species being compared. While there are many degrees of variation within each of these four possibilities, we offer this general framework and examples from the literature for each combination of extremes. (A) Loci with low sequence identity and low epigenetic similarity may represent lineage-specific loci and include non-orthologous regions. (B) A minority of orthologs demonstrate faster epi-mark divergence than sequence divergence. (C) Orthologs where the genome sequence is diverging faster than the epigenetic state represent loci that have experienced enhancer turnover. In addition, some marks are found in both fast and slowly evolving sequences, suggesting a mechanism for buffering genetic variation. (D) The majority of examples we can categorize exhibit both sequence conservation and epigenetic conservation. This is the status for most orthologs (inter- and intra-species) and represents the null hypothesis. Legend as in Figure 1; the intensity of shading of DNA strands represents the degree of sequence conservation ([26,30,32,34,75,76,105]).

elevated sequence conservation [31]. Unlike species-specific DNA methylated regions, human cortex-specific gains of acetylated H3K27 (H3K27ac) or H3K4me2 did not have concomitant accelerated sequence evolution compared to rhesus macaque or mouse brain cortex samples [33].

The first study to calculate epigenetic conservation explicitly in the context of sequence evolution analyzed a panel of epigenomic features in human, mouse, and pig ESCs [34]. The authors quantified epigenetic conservation between two species for a given epigenetic mark as the ratio

Box 1. Locus-Specific Example of Epigenome Evolution: the *c-FMS* Locus

Macrophage colony-stimulating factor receptor (c-FMS) expression marks hematopoietic commitment to the myeloid fate. Alternative first exons accompany transcripts in placental trophoblasts, and *c-FMS* is also an oncogene. Accordingly, *c-FMS* is subject to specific transcriptional regulation. The *c-FMS* loci in human and mouse have high sequence identity, especially at each alternative promoter and two intronic enhancers [26,102]. Despite high sequence conservation, *c-FMS* regulatory regions are bound by the same ensemble of transcription factors, but in different arrangements, along with some species-specific transcription factors [26]. However, the transcription factor ensembles recruit the same chromatin-remodeling factors in each species (Brg-1, HDAC) and drive the same transcriptional output in a cell-type and developmental-specific manner [26]. This example supports the hypothesis that evolutionarily conserved regulation may be driven by evolutionarily conserved regulatory element sequence and transcriptional programs, although there is interspecies variability in the execution of such a program.

of observed orthologous nucleotides conserved for that epi-mark over expectation. They then calculated the epigenetic conservation score for each modification over a range of binned PhyloP scores, a proxy for nucleotide substitution rate [35]. It was found that, regardless of species being compared, epigenetic score profiles fell into three distinct patterns. First, subsets of epi-marks (Polycomb-deposited H3K27me3 and promoter-associated H3K4me3) were conserved more often than expected at orthologous regions with low substitution rates, agreeing with the hypothesis that epigenetic conservation is correlated with genetic conservation. Second, three marks (DNA methylation, gene body-associated H3K36me3, and enhancer-associated H3K27ac) were enriched for conserved nucleotides over orthologous sequences with reduced substitution rate and sequences with an accelerated substitution rate in the human genome. The remainder of interrogated epi-marks (H3K9me3, H3K4me1, H3K4me2, and H2A.Z) had a uniform level of epigenetic conservation regardless of sequence evolution. Accordingly, fast-diverging orthologous sequences are more conserved than expected for DNA methylation, H3K36me3, and H3K27ac [34] (Figure 2), another example that sequence conservation is not required for conserved epi-mark status. The authors suggest that epi-conserved but genetically non-conserved regions may buffer against genetic mutations and provide functional stability to fast-evolving genome regions [34].

Sequence conservation is not always required for epi-conservation [36]. Instead, some epi-marks (DNA methylation, H3K36me3, and H3K27ac) may be conserved over orthologs whose sequence is under purifying selection. Understanding what features cause fast-evolving DNA fragments to undergo divergent epigenetic evolution is an important area of future study (Figure 3, Key Figure).

Epi-Mark Influence on Conserved (or Divergent) Gene Regulation

Epigenetic Conservation at Promoters

The promoter contains the regulatory DNA sequences surrounding a gene TSS and is responsible for transcription initiation. Epigenetic modifications at vertebrate promoters are well-studied; epi-marks common to active promoters are depicted in Figure 1.

DNA Methylation Conservation Status at Promoters

CpG islands (CGI) have long been recognized as non-methylated regions associated with protein-coding gene promoters [37]. Isolation of non-methylated DNA fragments is achieved via affinity purification with biotinylated CxxC (Bio-CAP), which preferentially binds to non-methylated DNA (Figure 1). Genome-wide Bio-CAP experiments followed by massively parallel sequencing in seven vertebrate genomes revealed that non-methylated islands (NMI) are a conserved feature of orthologous promoters, as well as of distal regulatory elements [38].

Methylation of CpGs in promoters is negatively correlated with gene expression [39,40], and some vertebrate CGIs are tissue-specifically methylated [38,41,42]. Array-based analysis of DNA methylation at ~27 000 CpG loci at proximal promoters in heart, liver, and kidney samples from human and chimp found that 18–26% of tissue-specific differentially methylated regions

Key Figure

Model for Building a Theory of Epigenome Evolution

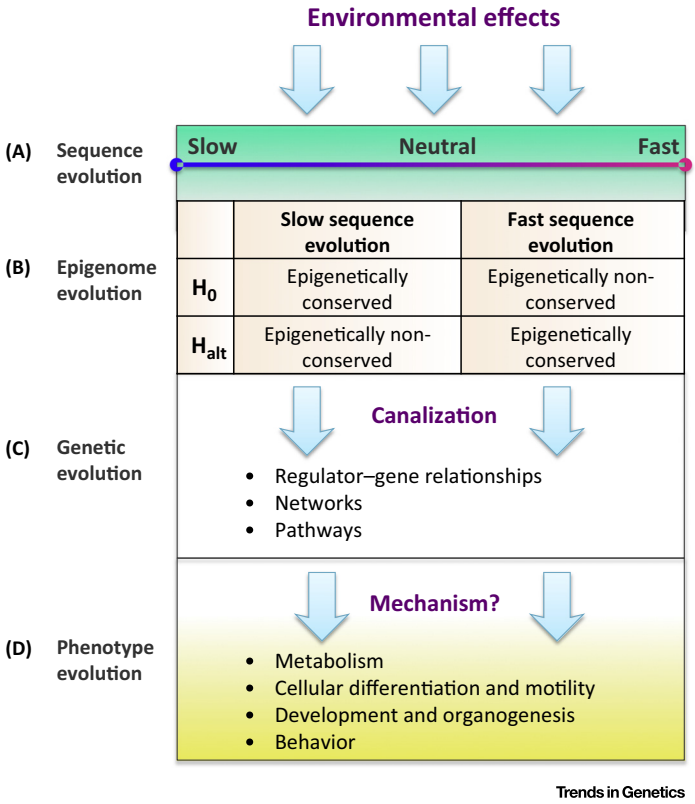


Figure 3. (A) Determining the rate of sequence evolution is now a straightforward process. (B) The expectation for epigenome evolution is different depending on the sequence evolution context. Completing this contingency table with specific examples is a challenge for the field. (C) Epigenetic gene regulation that is adaptive is genetically assimilated into the genome, codifying gene regulation and driving genetic evolution. (D) Genetic networks drive phenotypic evolution, all of which is motivated by environmental inputs. Abbreviations: H_0, null hypothesis; H_{alt}, alternative hypothesis.

(tsDMRs) were conserved between human and chimp (varying by tissue). Conserved tsDMRs were enriched for negative correlations between methylation level and the expression level of the associated gene (72% negative correlation values, regardless of species) [42]. In addition, promoters with conserved tsDMRs were enriched for genes annotated as 'developmental process' genes. Thus, epigenetically mediated tissue-specific regulation over core developmental genes tends to be conserved between human and chimp [42].

A similar study queried the DNA methylation status of ~326 000 probes shared between human, chimpanzees, bonobos, gorillas, and orangutan peripheral blood. Inter-species differentially methylated CpGs were depleted from proximal promoters and CpG islands (CGIs), suggesting broad conservation of promoter methylation status between primates [31]. However, there are promoters with inter-species differential DNA methylation, and these explained 12–18% of gene expression level differences between primates [42]. Thus species-differential DNA methylation at promoters can mediate species-differential gene expression. What sequence features underlie conserved or divergently methylated promoters?

CpG methylation status varies by CpG density, where denser CpG regions such as CGIs tend to have low levels of methylation, whereas regions of sparse or intermediate CpG density are variably methylated [43–45]. However, comparative analysis suggests that CpG density does not fully explain DNA methylation status at promoters [46] or predict DNA methylation divergence between paralogs [30,47]. Indeed, sequence analysis of experimentally determined NMIs revealed that the ratio of observed over expected CpGs and GC content of NMIs varies in a species-dependent manner [38]. Therefore, while NMIs are a common feature of vertebrate genomes and central to promoter regulation, the specific underlying sequence characteristics driving this conserved epigenetic feature may vary between species. Instead, TFBS motifs at methylation-determining regions [46] were found to be sufficient for proper methylation status at promoters [46–50]. In summary, while sequence features cannot fully predict DNA methylation status, transcription factors can alter DNA methylation patterns at promoters and drive promoter function. Comparative epigenome analysis incorporating multiple species will be necessary to understand how sequence evolution and motif turnover drive DNA methylation status at promoters (Figure 4).

Histone PTM Conservation

H3K4me3 defines active gene promoters [49]. Comparative analysis of ChIP for H3K4me3 and H3K27ac followed by sequencing (ChIP-seq) in liver samples from 20 mammals demonstrated that the basic regulatory landscape was very similar between species, including an average of ~12 500 active promoter elements [51]. Similarly, H3K4me3 enrichment was conserved in a cell type-specific manner between mouse and human, where ~80% of queried promoters were conserved in four homologous cell types [51]. Genome-wide, the magnitude of H3K4me3 promoter conservation increases in more closely related phylogenies: 16% of human liver promoters displayed conservation of H3K4me3 ChIP-seq signal across 20 mammals [51], while ~36% of orthologous promoters were conserved for H3K4me3 ChIP-seq between human, chimp, and rhesus macaque lymphoblastoid cell lines [52].

Species-differential promoter histone modification can also indicate species-specific gene expression [36]: up to 7% of differentially expressed genes between human and chimpanzee were explained by H3K4me3 distinctions [52]. These studies reveal that H3K4me3 and H3K27ac are features of highly expressed genes across mammals, and quantify how homologous cell types utilize orthologous genes for shared or species-specific functions. What sequence features drive conserved histone promoter marks, or mediate their turnover, remain to be investigated.

Gene Body Epi-Mark Conservation

The ‘gene body’ is the collection of introns and exons in the open reading frame of a gene. An archetypical vertebrate gene body epi-modifications is depicted in Figure 1.

Differential DNA Methylation over the Intron–Exon Junction

Seminal surveys of eukaryotic methylomes determined that CpG methylation over gene bodies is a conserved feature of eukaryotic genomes [52,53]. Internal exons typically display 6–20% elevated CG methylation compared to flanking introns [52], and gene body CpG methylation is conserved at 70–76% in human and chimp prefrontal cortex samples [54]. In both human and mouse genomes, recently duplicated genes retain conserved methylation patterns at gene body regions [47,55], suggesting that there is an overarching epigenetic mechanism that can identify duplicated fragments and properly methylate them. In most somatic tissues, high gene body methylation correlates with intermediate expression levels [53], with the exception of primate brain samples where gene body CpG methylation decreased linearly with increasing levels of gene expression [56].

Mammalian placenta are remarkable for their conserved global hypomethylation compared to somatic tissues (<66% genome-wide methylation level by MethylC-seq in human, rhesus

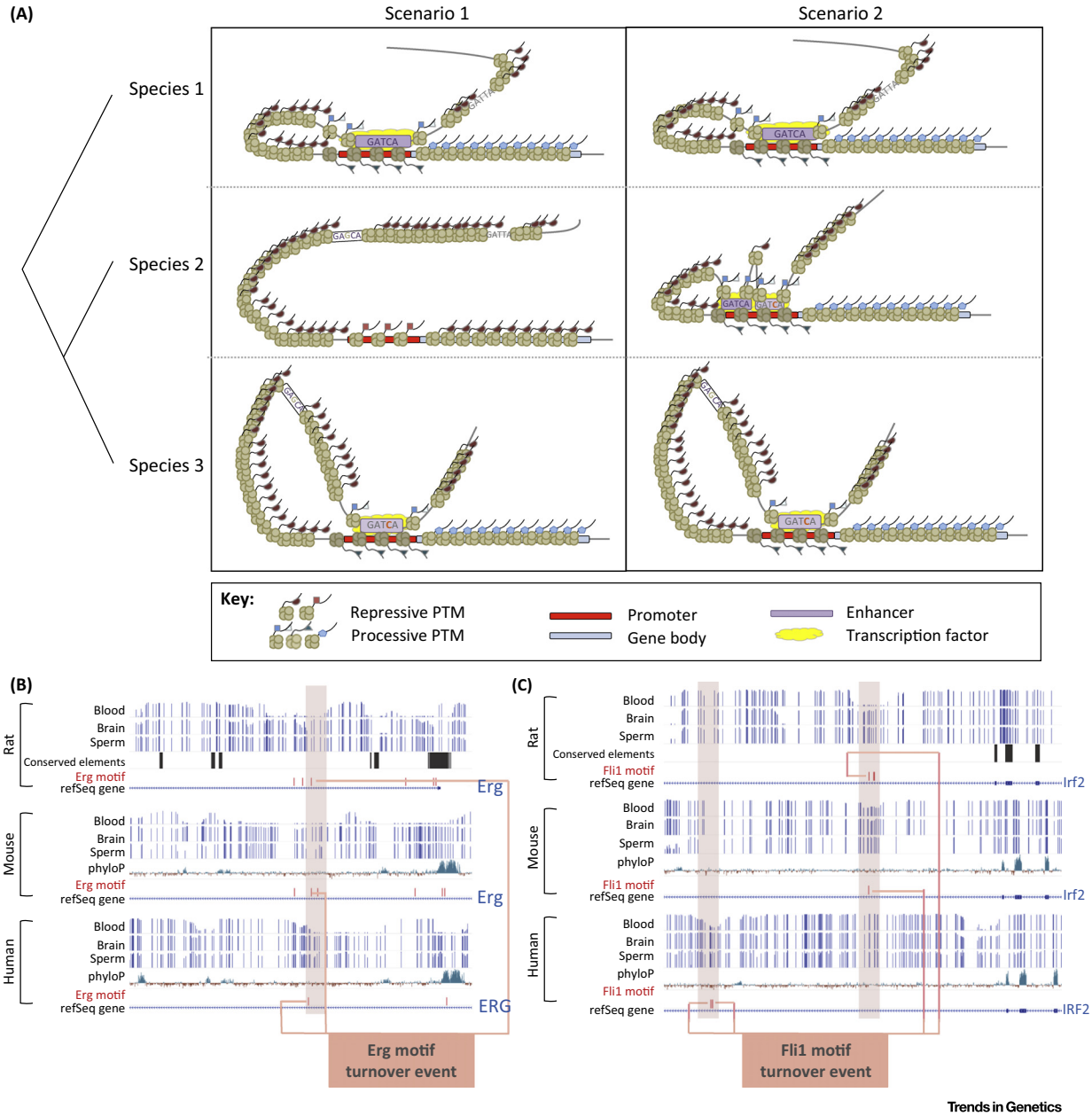


Figure 4. Transcription Factor Binding Site (TFBS) Turnover Models and Examples. Understanding TFBS turnover during evolution has been a non-trivial challenge. Because TFBS turnover is coupled with epigenetic changes, a null hypothesis that TFBS turnover is also associated with epigenetic evolution across species can be proposed. Although numerous examples of TFBS turnover have been documented, we propose two simple but powerful scenarios that can capture the process of TFBS turnover by comparing epigenetic signal across orthologous regions across species. The diagram illustrates an orthologous gene region across three species. This can be interpreted differently by varying the window size or synteny. The first scenario represents loss–gain TFBS turnover where species 1 had a TFBS but species 2 lost the TFBS by a single mutation in the binding site. However, in species 3, another genomic region was mutated to recover the lost TFBS and become the new enhancer. The second scenario is a competitive model where species 2 gained a mutation that generated another TFBS that competed with the species 1 enhancer. After selection or mutation, the species 1 enhancer is lost and the novel enhancer becomes the sole *cis*-regulator for the gene in species 3. This mechanism may mediate lineage-specific epigenetic marks [21,72,75,76,78] or conserve epigenetic features as in (B) and (C) [47]. (B) and (C) are representative examples of TFBS turnover events mediating a conserved tissue-specific DNA hypomethylated regions (pink shaded boxes) between rat, mouse, and human (adapted from J. Zhou, unpublished). Tracks in blue are single-CpG DNA methylation levels from the given tissue in each species. In (B) the Erg motif is found at the same position in rat and mouse, but is shifted by 84 bp in human. The Erg motif conserved the blood-specific DMR at this locus and is an example of scenario 1 depicted in (A). In (C), The Fli1 motif is in a slightly different position in the conserved blood-specific DMR in rat and mouse, but is absent in human. Instead the Fli1 motif is found in a nearby blood-specific DMR in human. Importantly, the conserved DMRs in these examples show low sequence conservation, providing evidence that conserved DNA hypomethylated regions do not depend on sequence conservation. Abbreviations: DMR, differentially methylated region; PTM, post-translational modification.

macaque, squirrel monkey, mouse, dog, horse, and cow placentas, as well as opossum extraembryonic membrane [57]). One exception to placental hypomethylation was that gene bodies displayed elevated methylation across all species. As in somatic tissues, high methylation over gene bodies in the placenta correlated with intermediate gene expression level, while genes with low gene body methylation were less likely to be expressed [57]. In addition, high placental gene body methylation was conserved across species over genes with similar gene ontologies, including genes involved in cell cycle, protein localization, and chromatin modification [57]. Gene body methylation is a conserved feature of eukaryotic genomes, and methylation level has a parabolic relationship to gene expression level in most eukaryotic somatic tissues and placenta, although the mechanistic links between genic DNA methylation and expression level are still unclear.

Histone PTM Marks over Exons

Exon-specific H3K36me3 modifications are conserved across eukaryotes [54,58] and are associated with exon inclusion [58]. Indeed, in exons, ChIP-seq signals of H3K36me3 and, to a lesser extent, H3K79me1, H4K20me1, and H2BK5me1, were found to increase as gene expression level increased [54]. However, histone modification enrichments could be a downstream result of differential nucleosome occupancy over exons and introns, where introns tend to be nucleosome-depleted. Emerging evidence for the role of H3K36me3 in gene splicing [59] and mismatch repair [60] may help to resolve the functional importance of H3K36me3 modification over exons.

Furthermore, colocalization of H3K27me3 and H3K36me3 over gene exons is associated with monoallelic gene expression, and this signal was conserved in human and mouse [61]. This signature is conserved over genes important for embryonic development and cell-surface protein genes. Notably, monoallelic expression and the corresponding epigenetic signatures of the genes were lineage-specific and maintained in differentiated tissues [61], suggesting that H3K36me3/H3K27me3 modifications may play a role in maintaining monoallelic expression and expression regulation in general [61].

Evolution of Epigenetic Regulation at Vertebrate Enhancers

Epigenetic modifications often have complimentary functions and are studied together to explore specific genetic elements. Enhancers display a characteristic histone modification profile of H3K4me1 and H3K27ac, usually in association with p300 [62,63], are usually hypomethylated [64], and are responsible for regulating cell type-appropriate gene expression [65]. While the characterization of novel enhancers is improving through advances in profiling techniques and computational models [66], our understanding of enhancer evolution is still unfolding.

Comparative studies analyzing histone PTM ChIP-seq signals between human and mouse developing heart [67], limb [36], and adipogenesis [68], and human and chimp cranial neural crest [69], and between developmental stages in distantly related zebrafish and medaka [70], reveal both shared and lineage-specific epi-marks. One study found that human neural crest cell (NCC) enhancers, defined by colocalization of H3K27ac and H3K4me3, showed strong enrichment of H3K27ac in the chicken orthologs and conserved TFAP2A binding in both species [71], suggesting these are conserved NCC enhancers. This study confirmed that several of these predicted enhancers have gene regulatory capabilities by reporter assays in zebrafish, and demonstrated that TFAP2A binding was necessary for specific enhancer activity [71]. Such multi-species analysis of cell type enhancer elements shows that histone modification can be a strong indicator of conserved, functional enhancers.

Leveraging epigenomic data from multiple organisms can identify species-specific enhancer elements [36,67,72,73]. DNase I footprints are identified from DNase I-seq datasets where the

cleavage pattern of DNase I digested fragments is abrogated by a DNA-binding protein, such as a transcription factor, occupying the DNA [74]. DNase I footprinting analyses in human and mouse revealed that, although 65% of DNase I footprints in mouse have an orthologous sequence in human, only 22% of those orthologs also showed a DNase I footprinting signal, suggesting that there has been a large-scale turnover of transcription factor binding since the human–rodent split [72,75–78]. An assessment of DNase I hypersensitive sites (DHSs) in human, chimp, and macaque skin fibroblasts and lymphoblastoid cell lines showed that most DHSs are conserved across species because pairwise comparison of genome-wide DHS signals revealed that these were highly correlated [79]. However, several hundred DHSs were gained or lost in each of the human and chimp lineages, particularly at distal enhancers and introns [79].

Conversely, regulatory element conservation decreases when the number of epigenomes being compared is increased. ChIP-seq in liver samples from 20 mammals showed that, although enhancers were more common than promoters, only 1% of human liver enhancers had conserved H3K27ac signal at the orthologous sequence in at least 10 other mammalian genomes [51].

Species-specific enhancers are clearly common, but are they functional? Human DHS gains in skin and lymphoblastoid cells significantly overlapped with ChIP-seq signals for enhancer-associated chromatin marks H3K4me1 (~80% overlap), H3K4me2 (~80%), and H3K27ac (~70%) [79]. Lineage-specific epi-marks were enriched near tissue-relevant genes, and genes associated with lineage-specific marks were discordantly expressed between species [36,51,68,69]. Although more conserved than the genomic background, human limb-specific enhancers were found to have less sequence conservation than limb enhancers shared with rhesus macaque and mouse [36]. In addition, human liver DHS gains/losses had stronger signal of positive selection on the human lineage, suggesting these regions are likely functional and may contribute to specific-specific gene regulation [79]. The high rate of lineage-specific enhancer turnover may be driven by **transcription factor binding site (TFBS) turnover** (Figure 4; see TFBS Turnover as a Mechanism for Epigenome Evolution) [67,68,79].

Transcription Factor Binding Occupancy at Orthologs

With the publication of the Mouse ENCODE Project, genome wide large-scale comparative analysis between mouse and human transcription factor ChIP-seq data is now available [80]. Because a comprehensive review of the genetics of transcription factor occupancy across species has been published elsewhere [73], in this review we highlight what has been learned about the epigenetic context of transcription factor binding to human/mouse orthologous regulatory regions from the Mouse ENCODE Project.

Cis-regulatory sequences in mouse are enriched for conserved sequences: ~67% of both DNase I hypersensitivity sites and transcription factor ChIP-seq peaks had homologs in human, while ~79% of both chromatin-based promoter and enhancer predictions had homologs in human [80]. However, a smaller fraction of the human orthologs of predicted regulatory sequences in mouse were also predicted to be promoters (44%) or enhancers (40%) in human [80], suggesting that regulatory element conservation does not always track with sequence similarity [73,81]. How did species-specific regulatory elements evolve? The authors found that 89% of histone-defined mouse-specific promoters and 85% of mouse-specific enhancers overlap transposable elements (TEs) or mobile elements, and were enriched for specific TE classes [80], suggesting that DNA derived from transposable elements may be responsible for a large fraction of species-specific gene regulation [76].

To examine transcription factor ChIP-seq binding peaks in a cell type-specific manner, the binding profiles of 32 transcription factors in two human and mouse homologous cell types:

erythroid progenitors (mouse MEL; human K562) and lymphoblastoid cells (mouse CH12; human GM12878) were examined [75]. The factors queried included Pol2, CTCF, and other general and cell type-specific transcription activators. Conservation of transcription factor occupancy for an orthologous site varied in a factor-specific manner, and conservation was highest at proximal promoter regions (even after controlling for elevated sequence conservation at promoters), with the exception of CTCF [75]. Epigenetic modification mimicked transcription factor binding occupancy across species: DNA methylation levels were low in both species over orthologs with occupancy-conserved binding, but DNA methylation increased over unbound orthologs. In aggregate, bound fragments show elevated evolutionary constraint, but ~50% of bound regions in one species were not alignable in the other, representing species-specific binding events that may be mediated in some instances by TEs [75].

In the vertebrate genome, repetitive sequences and TEs contribute ~6–60% of total sequence content [82]. However, TE sequences are thought to be silenced through epigenetic defense mechanisms because transposition events can be deleterious [83]. Interestingly, the dynamic epigenetic silencing of TEs during development is conserved among vertebrates, although numerous TEs are known to be species-specific. Furthermore, TEs are hypothesized to shape gene regulatory networks through exaptation [84]. Exaptation describes the process where the TEs evolved to acquire new function in the genome, such as novel TFBSs, that provided some fitness benefits in the host [84]. There are two models of TE exaptation: (i) a surplus of TE insertions in the genome provided raw sequence material that can be mutated into novel TFBSs, and (ii) TEs with a functional TFBS transposed throughout the genome until a functional gain that led to a fitness benefit and fixation [85].

Recent work revealed that TEs contributed 2–40% of transcription factor binding events in human or mouse, depending on the factor and cell type [76]. Nevertheless, only 2% of human TE-derived transcription factor binding sites and 1% of mouse TE-derived sites were occupied by the same transcription factor at a syntenic site in the opposite genome. Furthermore, 99% of human and 98% of mouse TE-derived binding sites were species-specific, suggesting either the host TE amplified after the primate–rodent split or that the TE ancestor accumulated too many mutations to be recognized as a TE sequence in the other genome [76] (Figure 2). Regardless, this comparative analysis supports the hypothesis that TEs may rewire gene regulatory networks in a species-specific manner [19,20].

TFBS Turnover as a Mechanism for Epigenome Evolution

TFBS turnover can explain species-specific transcription factor binding events [75,76,78,86,87]. For example, across 4000 orthologous promoters between mouse and human, 41–89% of liver transcription factor binding locations were species-specific [77], suggesting a high amount of TFBS turnover since the last mouse/human common ancestor [72,88].

Formation of novel TFBSs can disrupt and shift methylation patterns in the promoter region [46,48,89]. For example, one study described methylation-determining regions that direct the DNA methylation status of promoter-proximal CpGs during differentiation [46]. TFBS sites, including SP1, CTCF, and Rfx, were required for proper methylation [46]. Moreover, RE1 silencing transcription factor (REST)-binding regions of low-methylation (LMRs) showed increased DNA methylation in *Rest* knockout ESCs, indicating that REST is required for proper demethylation of LMRs [49]. Thus, evidence is mounting that DNA sequence polymorphisms in TFBSs may modulate DNA methylation status.

Accordingly, TFBS turnover events have been found to explain paralog- or lineage-specific differential DNA methylation [30,75,76,90,91], DNase hypersensitivity [79], histone PTMs [67,69], and transcription factor binding events [21,72,75,76,92]. Comparison of orthologous

CpGs in the primate lineage revealed that human-specific DMRs genome-wide were enriched for nucleotide substitutions in TFBSs, suggesting a close relationship between TFBSs and DNA methylation patterns during human evolution [32]. Similarly, motifs for chromatin regulators or transcription factors associated with a particular chromatin state (such as SP1 and CTCF, respectively) were enriched in epigenetically divergent paralogs [30]. Analysis of binding sites of pluripotency transcription factors in human, mouse, and pig ESCs demonstrated that inter-species epigenetic differences explain species-differential binding and expression better than sequence differences [34]. Overall, examination of binding-site turnover events and their epigenetic context supports the hypothesis that DNA sequence changes in the form of TFBS turnover events drive epigenetic variation that may regulate gene expression (Figure 4).

Specific TFBS motifs may also mediate epigenome conservation [92]. Preservation of DNA methylation patterns over duplicated genes was associated with the SP1 motif at paralogous promoters [47]. TFBSs also mark regulatory elements with DNA methylation (J. Zhou, unpublished) or histone modification conservation between species [51] (Figure 4B). In each case, TFBS turnover might be a mechanism for canalization, codifying epigenetic modifications into genetic knowledge and ensuring robustness of a phenotype [91,93] (Figure 3).

Concluding Remarks

To efficiently benefit from comparative epigenome research, a model of epigenome evolution needs to be established. Epigenome analysis is a quickly maturing field, and the combination of epigenomics and computational modeling with classic technologies can make inroads on previously intractable questions of epigenetic gene regulation. Nevertheless, comparative epigenomics faces some challenges and limitations (Box 2).

To build a model of epigenome evolution, the field must first answer basic questions about how epigenetic conservation relates to sequence evolution (Figure 3). For example, what types of epi-modifications occur at slowly evolving DNA sequences compared to fast evolving sequences? At slowly evolving sequences, we suggest the null hypothesis is that orthologs should display conserved epigenetic modifications; the alternative hypothesis is that DNA-conserved orthologs have epi-divergent status. In this case support for the alternative hypothesis is evidence for regulatory innovation driven by epigenetic novelty. At quickly evolving DNA loci the hypotheses

Box 2. Challenges and Limitations for Comparative Epigenomics

Because epigenetic status varies with cell type, matching homologous tissues or cell types between species is required for rigorous epigenssome comparison. However, matching homologous cell types between species is a non-trivial task [6,103], especially when developmental stage and environment may also need to be matched. One complication is how to determine identity by descent when differentiated cell types must be specified every generation [104]. This might be achieved by comparative molecular cell biology, as has been shown in the case of retinal cell evolution [103]. The comparative molecular biology approach analyzes expression of orthologous genes to identify homologous cell types. However, the issue is made more complex by the realization that homologous genes may not direct the development of homologous structures [104]. Instead, it has been proposed that conserved gene regulatory networks (GRNs) may control the development of homologous structures, and these GRNs may comprise non-orthologous genes [104]. By taking into account the nuances involved in determining biological homology when designing experiments [6], comparative epigenomics may determine how conserved (or species-specific) epigenomic features of homologous cell types regulate GRNs.

Between distantly related species, establishing homology becomes increasingly challenging. The evolutionary history we can recover using comparative epigenomics is therefore limited. Other technical challenges remain, and these include variable genome build qualities and the accuracy of multiple genome sequence alignments [18]. Because all next-generation sequencing (NGS)-based assays are subject to batch effects, systematic biases in experimental design should be limited or corrected for as much as possible because combination with the technical limitations of inter-species analysis could hinder the interpretability of results. Lastly, given the challenges for multi-species sequence alignment, how do we go about aligning the epigenome? Nucleotide bases are the units of the genome sequence; what are the units of the epigenome?

Outstanding Questions

Hybrid cell culture systems can be used to study lineage-specific epigenome features. Cells harboring chromosomes from different species may help to disentangle the environmental, *trans*, or *cis* effects on epigenome status.

Genomic editing technologies make genome editing feasible in any species, and can identify drivers of conserved and divergent epigenetic programs. Targets in homologous cell models may be altered to understand how the *cis*-landscape impacts on epigenome state in different species. Similarly, analyzing epigenome status after knockout of tissue-specific transcription factors or chromatin modifiers may point to epigenetic effects that are conserved or species-specific, revealing mechanisms of the evolution of complex tissues and organs.

Applying single-cell technologies to comparative epigenomics can identify cell-type specific and temporal differences at higher resolution, elucidating how epi-modifications depend on sequence in an allele-specific manner.

While the determinants of nucleosome positioning in eukaryotic cells are similar, quantitative comparative studies of nucleosome positioning is best explored in yeast. Some comparative analyses between phyla exist, but comparative nucleosome positioning in vertebrate genomes is a promising an area for future research.

Hi-C technology permits comparative analysis of chromatin architecture and nuclear territories. 3D fluorescent *in situ* hybridization experiments have demonstrated that chromosome territories in the nucleus are largely conserved across primates. Comparative analyses of chromosome conformation can reveal how genome organization evolved and how nuclear architecture contributes to gene regulation.

Both tumor cells and iPSC undergo epigenetic reprogramming. Oncogenic transformation may model epigenome evolution on an accelerated timescale. iPSC–cancer cell comparative epigenomics can expose how epigenetic misregulation contributes to oncogenesis.

are flipped: the null hypothesis is that epi-marks will be different, but the alternative hypothesis is that they would be conserved. Evidence of the alternative hypothesis in this case would be a signature of epigenetic buffering [34,91].

Few studies have quantified the magnitude of **epigenetic divergence** or conservation with respect to sequence evolution. As a result, it is not yet clear how sequence evolution influences epigenome evolution (Figure 2). To ameliorate this knowledge gap, the table in the center of the proposed framework (Figure 3) should be populated with comparative epigenome studies that take into account sequence evolution as well as inter-species epigenetic differences. First, conservation of three epi-marks (DNA methylation, H3K36me3, and H3K27ac) in mammalian ESCs is independent of DNA sequence evolution [34], and the regulatory elements identified in Xiao *et al.* [34] fulfill both the null hypothesis (H_0) for slowly evolving sequences and the alternative hypothesis (H_{alt}) for fast evolving sequences. By contrast, ample evidence exists that gene body epi-marks are conserved and, because most gene bodies are slowly evolving [94], the conservation of gene body epigenetic regulation is evidence for the null hypothesis of epigenome evolution over slowly evolving sequences (Figure 3).

Beyond evidence for each hypothesis, this framework prompts other questions such as what features distinguish divergent versus conserved epi-marks at fast evolving DNA? How and at what rate can a regulatory element transition from a slow DNA-evolving/epigenetically divergent status to a fast DNA-evolving/epigenetically conserved status? What mechanisms mediate genetic assimilation of stable epigenetic regulatory architectures to be genetically encoded?

We view genetic evolution (the evolution of networks and pathways) as subsequent to epigenetic evolution. We note that comparative analysis of genetic networks has found that conservation of regulatory motifs and network topologies between mouse and human is much higher than conservation of individual transcription factor binding sites [95]. Thus we suggest that evolution at the epigenome level is what mediates the flexibility of transcription factor binding sites and thus the innovation or stability of gene regulatory networks.

In the Outstanding Questions, key areas of research are proposed that will help to fill in the proposed framework with empirical data. Lastly, rigorous mathematical models must accompany a mature model of epigenome evolution that quantitatively assesses epigenome conservation or divergence based on data, in a manner similar to the Jukes–Cantor model for sequence evolution.

References

1. Montavon, T. and Duboule, D. (2013) Chromatin organization and global regulation of Hox gene clusters. *Philos. Trans. R. Soc. Lond. B: Biol. Sci.* 368, 20120367
2. Lukas, J. *et al.* (2011) More than just a focus: the chromatin response to DNA damage and its role in genome integrity maintenance. *Nat. Cell Biol.* 13, 1161–1169
3. Okamoto, I. *et al.* (2011) Eutherian mammals use diverse strategies to initiate X-chromosome inactivation during development. *Nature* 472, 370–374
4. Chow, J. and Heard, E. (2009) X inactivation and the complexities of silencing a sex chromosome. *Curr. Opin. Cell Biol.* 21, 359–366
5. Leeb, M. and Wutz, A. (2012) Establishment of epigenetic patterns in development. *Chromosoma* 121, 251–262
6. Wagner, G.P. (1989) The biological homology concept. *Annu. Rev. Ecol. Syst.* 20, 51–69
7. Gerstein, M.B. *et al.* (2010) Integrative analysis of the *Caenorhabditis elegans* genome by the modENCODE project. *Science* 330, 1775–1787
8. Consortium, modENCODE. *et al.* (2010) Identification of functional elements and regulatory circuits by *Drosophila* modENCODE. *Science* 330, 1787–1797
9. Consortium, R.E. *et al.* (2015) Integrative analysis of 111 reference human epigenomes. *Nature* 518, 317–330
10. Mouse ENCODE Consortium *et al.* (2012) An encyclopedia of mouse DNA elements (Mouse ENCODE). *Genome Biol.* 13, 418
11. Kimura, M. (1968) Evolutionary rate at the molecular level. *Nature* 217, 624–626
12. Pennacchio, L.A. *et al.* (2006) In vivo enhancer analysis of human conserved non-coding sequences. *Nature* 444, 499–502
13. Clarke, S.L. *et al.* (2012) Human developmental enhancers conserved between deuterostomes and protostomes. *PLoS Genet.* 8, e1002852
14. Margulies, E.H. *et al.* (2007) Analyses of deep mammalian sequence alignments and constraint predictions for 1% of the human genome. *Genome Res.* 17, 760–774
15. Cooper, G.M. and Brown, C.D. (2008) Qualifying the relationship between sequence conservation and molecular function. *Genome Res.* 18, 201–205
16. King, M.C. and Wilson, A.C. (1975) Evolution at two levels in humans and chimpanzees. *Science* 188, 107–116
17. Wray, G.A. (2007) The evolutionary significance of cis-regulatory mutations. *Nat. Rev. Genet.* 8, 206–216

18. Chen, X. and Tompa, M. (2010) Comparative assessment of methods for aligning multiple genome sequences. *Nat. Biotechnol.* 28, 567–572
19. Britten, R.J. and Davidson, E.H. (1969) Gene regulation for higher cells: a theory. *Science* 165, 349–357
20. McClintock, B. (1956) Controlling elements and the gene. *Cold Spring Harb. Symp. Quant. Biol.* 21, 197–216
21. Kunarso, G. *et al.* (2010) Transposable elements have rewired the core regulatory network of human embryonic stem cells. *Nat. Genet.* 42, 631–634
22. Blow, M.J. *et al.* (2010) ChIP-Seq identification of weakly conserved heart enhancers. *Nat. Genet.* 42, 806–810
23. Dogan, N. *et al.* (2015) Occupancy by key transcription factors is a more accurate predictor of enhancer activity than histone modifications or chromatin accessibility. *Epigenetics Chromatin* 8, 16
24. Koide, T. *et al.* (1994) Comparative analysis of Igf-2/H19 imprinted domain: identification of a highly conserved intergenic DNase I hypersensitive region. *Genomics* 24, 1–8
25. Hardison, R. *et al.* (1997) Locus control regions of mammalian β-globin gene clusters: combining phylogenetic analyses and experimental results to gain functional insights. *Gene* 205, 73–94
26. Follows, G.A. *et al.* (2003) Differential transcription factor occupancy but evolutionarily conserved chromatin features at the human and mouse M-CSF (CSF-1) receptor loci. *Nucleic Acids Res.* 31, 5805–5816
27. Arney, K.L. *et al.* (2006) The human and mouse H19 imprinting control regions harbor an evolutionarily conserved silencer element that functions on transgenes in *Drosophila*. *Dev. Genes Evol.* 216, 811–819
28. Smits, G. *et al.* (2008) Conservation of the H19 noncoding RNA and H19–IGF2 imprinting mechanism in therians. *Nat. Genet.* 40, 971–976
29. Killian, J.K. *et al.* (2000) M6P/IGF2R imprinting evolution in mammals. *Mol. Cell* 5, 707–716
30. Prendergast, J.G.D. *et al.* (2014) Sequence-level mechanisms of human epigenome evolution. *Genome Biol. Evol.* 6, 1758–1771
31. Hernando-Herraez, I. *et al.* (2013) Dynamics of DNA methylation in recent human and great ape evolution. *PLoS Genet.* 9, e1003763–e1003812
32. Hernando-Herraez, I. *et al.* (2015) The interplay between DNA methylation and sequence divergence in recent human evolution. *Nucleic Acids Res.* 43, 8204–8214
33. Reilly, S.K. *et al.* (2015) Evolutionary changes in promoter and enhancer activity during human corticogenesis. *Science* 347, 1155–1159
34. Xiao, S. *et al.* (2012) Comparative epigenomic annotation of regulatory DNA. *Cell* 149, 1381–1392
35. Siepel, A. *et al.* (2006) New methods for detecting lineage-specific selection. In *Research in Computational Molecular Biology* (Vol. 3909), pp. 190–205, Springer Berlin Heidelberg
36. Cotney, J. *et al.* (2013) The evolution of lineage-specific regulatory activities in the human embryonic limb. *Cell* 154, 185–196
37. Aïssani, B. and Bernardi, G. (1991) CpG islands: features and distribution in the genomes of vertebrates. *Gene* 106, 173–183
38. Long, H.K. *et al.* (2013) Epigenetic conservation at gene regulatory elements revealed by non-methylated DNA profiling in seven vertebrates. *Elife* 2, e00348
39. Keshet, I. *et al.* (1985) Effect of regional DNA methylation on gene expression. *Proc. Natl. Acad. Sci. U.S.A.* 82, 2560–2564
40. Feenastr, A. *et al.* (1986) In vitro methylation inhibits the promotor activity of a cloned intracisternal A-particle LTR. *Nucleic Acids Res.* 14, 4343–4352
41. Illingworth, R. *et al.* (2008) A novel CpG island set identifies tissue-specific methylation at developmental gene loci. *PLoS Biol.* 6, e22
42. Pai, A.A. *et al.* (2011) A genome-wide study of DNA methylation patterns and gene expression levels in multiple human and chimpanzee tissues. *PLoS Genet.* 7, e1001316
43. Bird, A. (2002) DNA methylation patterns and epigenetic memory. *Genes Dev.* 16, 6–21
44. Elliott, G. *et al.* (2015) Intermediate DNA methylation is a conserved signature of genome regulation. *Nat. Commun.* 6, 6363
45. Suzuki, M.M. and Bird, A. (2008) DNA methylation landscapes: provocative insights from epigenomics. *Nat. Rev. Genet.* 9, 465–476
46. Lienert, F. *et al.* (2011) Identification of genetic elements that autonomously determine DNA methylation states. *Nat. Genet.* 43, 1091–1097
47. Keller, T.E. and Yi, S.V. (2014) DNA methylation and evolution of duplicate genes. *Proc. Natl. Acad. Sci. U.S.A.* 111, 5932–5937
48. Macleod, D. *et al.* (1994) Sp1 sites in the mouse Aprt gene promoter are required to prevent methylation of the CpG island. *Genes Dev.* 8, 2282–2292
49. Feldmann, A. *et al.* (2013) Transcription factor occupancy can mediate active turnover of DNA methylation at regulatory regions. *PLoS Genet.* 9, e1003994
50. Schübeler, D. (2015) Function and information content of DNA methylation. *Nature* 517, 321–326
51. Villar, D. *et al.* (2015) Enhancer evolution across 20 mammalian species. *Cell* 160, 554–566
52. Feng, S. *et al.* (2010) Conservation and divergence of methylation patterning in plants and animals. *Proc. Natl. Acad. Sci. U.S.A.* 107, 8689–8694
53. Zemach, A. *et al.* (2010) Genome-wide evolutionary analysis of eukaryotic DNA methylation. *Science* 328, 916–919
54. Schwartz, S. *et al.* (2009) Chromatin organization marks exon–intron structure. *Nat. Struct. Mol. Biol.* 16, 990–995
55. Chang, A.Y-F. and Liao, B-Y. (2012) DNA methylation rebalances gene dosage after mammalian gene duplications. *Mol. Biol. Evol.* 29, 133–144
56. Zeng, J. *et al.* (2012) Divergent whole-genome methylation maps of human and chimpanzee brains reveal epigenetic basis of human regulatory evolution. *Am. J. Hum. Genet.* 91, 455–465
57. Schroeder, D.I. *et al.* (2015) Early developmental and evolutionary origins of gene body DNA methylation patterns in mammalian placentas. *PLoS Genet.* 11, e1005442
58. Kolasinska-Zwierz, P. *et al.* (2009) Differential chromatin marking of introns and expressed exons by H3K36me3. *Nat. Genet.* 41, 376–381
59. Luco, R.F. *et al.* (2010) Regulation of alternative splicing by histone modifications. *Science* 327, 996–1000
60. Li, F. *et al.* (2016) Regulation of mismatch repair by histone code and posttranslational modifications in eukaryotic cells. *DNA Repair* 38, 68–74
61. Nag, A. *et al.* (2015) Chromatin signature identifies monoallelic gene expression across mammalian cell types. *G3* 5, 1713–1720
62. Heintzman, N.D. *et al.* (2009) Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature* 459, 108–112
63. Zhu, Y. *et al.* (2013) Predicting enhancer transcription and activity from chromatin modifications. *Nucleic Acids Res.* 41, 10032–10043
64. Stadler, M.B. *et al.* (2011) DNA-binding factors shape the mouse methylome at distal regulatory regions. *Nature* 480, 490–495
65. Rubinstein, M. and de Souza, F.S.J. (2013) Evolution of transcriptional enhancers and animal diversity. *Philos. Trans. R. Soc. Lond. B: Biol. Sci.* 368, 20130017
66. Wilczynski, B. *et al.* (2012) Predicting spatial and temporal gene expression using an integrative model of transcription factor occupancy and chromatin state. *PLoS Comput. Biol.* 8, e1002798
67. Hsu, C-H. and Ovcharenko, I. (2013) Effects of gene regulatory reprogramming on gene expression in human and mouse developing hearts. *Philos. Trans. R. Soc. Lond. B: Biol. Sci.* 368, 20120366
68. Mikkelsen, T.S. *et al.* (2010) Comparative epigenomic analysis of murine and human adipogenesis. *Cell* 143, 156–169
69. Prescott, S.L. *et al.* (2015) Enhancer divergence and cis-regulatory evolution in the human and chimp neural crest. *Cell* 163, 68–83
70. Tena, J.J. *et al.* (2014) Comparative epigenomics in distantly related teleost species identifies conserved cis-regulatory nodes active during the vertebrate phylotypic period. *Genome Res.* 24, 1075–1085

71. Rada-Iglesias, A. *et al.* (2012) Epigenomic annotation of enhancers predicts transcriptional regulators of human neural crest. *Cell Stem Cell* 11, 633–648
72. Yokoyama, K.D. *et al.* (2014) Tracing the evolution of lineage-specific transcription factor binding sites in a birth–death framework. *PLoS Comput. Biol.* 10, e1003771
73. Stefflova, K. *et al.* (2013) Cooperativity and rapid evolution of cobound transcription factors in closely related mammals. *Cell* 154, 530–540
74. Stamatoyannopoulos, J.A. *et al.* (1995) NF-E2 and GATA binding motifs are required for the formation of DNase I hypersensitive site 4 of the human beta-globin locus control region. *EMBO J.* 14, 106–116
75. Cheng, Y. *et al.* (2014) Principles of regulatory information conservation between mouse and human. *Nature* 515, 371–375
76. Sundaram, V. *et al.* (2014) Widespread contribution of transposable elements to the innovation of gene regulatory networks. *Genome Res.* 24, 1963–1976
77. Odom, D.T. *et al.* (2007) Tissue-specific transcriptional regulation has diverged significantly between human and mouse. *Nat. Genet.* 39, 730–732
78. Schmidt, D. *et al.* (2010) Five-vertebrate ChIP-seq reveals the evolutionary dynamics of transcription factor binding. *Science* 328, 1036–1040
79. Shibata, Y. *et al.* (2012) Extensive evolutionary changes in regulatory element activity during human origins are associated with altered gene expression and positive selection. *PLoS Genet.* 8, e1002789
80. Yue, F. *et al.* (2014) A comparative encyclopedia of DNA elements in the mouse genome. *Nature* 515, 355–364
81. Bernstein, B.E. *et al.* (2005) Genomic maps and comparative analysis of histone modifications in human and mouse. *Cell* 120, 169–181
82. Chalopin, D. *et al.* (2015) Comparative analysis of transposable elements highlights mobilome diversity and evolution in vertebrates. *Genome Biol. Evol.* 7, 567–580
83. Bejerano, G. *et al.* (2006) A distal enhancer and an ultraconserved exon are derived from a novel retroposon. *Nature* 441, 87–90
84. de Souza, F.S.J. *et al.* (2013) Exaptation of transposable elements into novel cis-regulatory elements: is the evidence always strong? *Mol. Biol. Evol.* 30, 1239–1251
85. Feschotte, C. (2008) Transposable elements and the evolution of regulatory networks. *Nat. Rev. Genet.* 9, 397–405
86. Dermitzakis, E.T. and Clark, A.G. (2002) Evolution of transcription factor binding sites in mammalian gene regulatory regions: conservation and turnover. *Mol. Biol. Evol.* 19, 1114–1121
87. Frith, M.C. (2006) Evolutionary turnover of mammalian transcription start sites. *Genome Res.* 16, 713–722
88. Stergachis, A.B. *et al.* (2014) Conservation of trans-acting circuitry during mammalian regulatory evolution. *Nature* 515, 365–370
89. Brandeis, M. *et al.* (1994) Sp1 elements protect a CpG island from de novo methylation. *Nature* 371, 435–438
90. Irizarry, R.A. *et al.* (2009) The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. *Nat. Genet.* 41, 178–186
91. Feinberg, A.P. and Irizarry, R.A. (2010) Colloquium paper: stochastic epigenetic variation as a driving force of development, evolutionary adaptation, and disease. *Proc. Natl. Acad. Sci. U.S.A.* 107, 1757–1764
92. Göke, J. *et al.* (2011) Combinatorial binding in human and mouse embryonic stem cells identifies conserved enhancers active in early embryonic development. *PLoS Comput. Biol.* 7, e1002304
93. Waddington, C.H. (1942) Canalization of development and the inheritance of acquired characters. *Nature* 150, 563–565
94. Lander, E.S. *et al.* (2001) Initial sequencing and analysis of the human genome. *Nature* 409, 860–921
95. Stergachis, A.B. *et al.* (2013) Developmental fate and cellular maturity encoded in human regulatory DNA landscapes. *Cell* 154, 888–903
96. Frommer, M. *et al.* (1992) A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. *Proc. Natl. Acad. Sci. U.S.A.* 89, 1827–1831
97. Meissner, A. *et al.* (2005) Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis. *Nucleic Acids Res.* 33, 5868–5877
98. Maunakea, A.K. *et al.* (2010) Conserved role of intragenic DNA methylation in regulating alternative promoters. *Nature* 466, 253–257
99. Blackledge, N.P. *et al.* (2012) Bio-CAP: a versatile and highly sensitive technique to purify and characterise regions of non-methylated DNA. *Nucleic Acids Res.* 40, e32
100. Barski, A. *et al.* (2007) High-resolution profiling of histone methylations in the human genome. *Cell* 129, 823–837
101. Thurman, R.E. *et al.* (2012) The accessible chromatin landscape of the human genome. *Nature* 489, 75–82
102. Tagoh, H. *et al.* (2002) Transcription factor complex formation and chromatin fine structure alterations at the murine c-Fms (CSF-1 receptor) locus during maturation of myeloid precursor cells. *Genes Dev.* 16, 1721–1737
103. Arendt, D. (2003) Evolution of eyes and photoreceptor cell types. *Int. J. Dev. Biol.* 47, 563–571
104. Wagner, G.P. (2007) The developmental genetics of homology. *Nat. Rev. Genet.* 8, 473–479
105. Chambers, E.V. *et al.* (2013) Divergence of mammalian higher order chromatin structure is associated with developmental loci. *PLoS Comput. Biol.* 9, e1003017