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.