Genomics 108 (2016) 115–125

Contents lists available at ScienceDirect

Genomics

journal homepage: www.elsevier.com/locate/ygeno

Review

Insights in human epigenomic dynamics through comparative primate analysis

Christopher G. Bell

Epigenomic Medicine, Biological Sciences, University of Southampton, United Kingdom
MRC Lifecourse Epidemiology Unit, Southampton General Hospital, University of Southampton, United Kingdom
Institute of Developmental Sciences, Southampton General Hospital, University of Southampton, Southampton, United Kingdom

ARTICLE INFO

Article history:
Received 3 April 2016
Received in revised form 3 August 2016
Accepted 29 September 2016
Available online 1 October 2016

Keywords:
Epigenomics
Epigenetics
Comparative genomics
DNA methylation
Chromatin
Histone modifications

ABSTRACT

Epigenomic analysis gives a molecular insight into cell-specific genomic activity. It provides a detailed functional plan to dissect an organism, tissue by tissue. Therefore comparative epigenomics may increase understanding of human-acquired traits, by revealing regulatory changes in systems such as the neurological, musculoskeletal, and immunological.

Enhancer loci evolve fast by hijacking elements from other tissues or rewiring and amplifying existing units for human-specific function. Promoters by contrast often require a CpG dense genetic infrastructure. Specific interplay occurs between the two, but also a shared modality of function, with coordination from global chromatin-modifying enzymes. Changes in specific transcription factor binding sites also facilitate the local epigenetic state. In the case of CTCF, these may further influence 3-dimensional structure and interaction.

How these mechanistic units are modulated between tissue and species enables more comprehensive understanding of human processes and pathology. With this information, precise therapeutic targeting of these epigenetic modifications may become possible.

Contents

http://dx.doi.org/10.1016/j.ygeno.2016.09.003
0888-7543/

1. Introduction

Our understanding of the evolution of our species, *Homo sapiens*, will be transformed in the coming decades [1]. Genomic tools are now able to uncover molecular mechanisms not only comparatively across primate species [2], but also to peer into the past as demonstrated by the recent insights into Neanderthal, Denisovan, and still other unknown extinct hominins [3,4]. These ancient DNA studies reveal that incorporated or 'introgressed' archaic DNA is found within the non-African modern human populations of today. This has both potential adaptive and disease susceptibility legacies [5–7], for example, evidence that the Denisovan genetic inheritance within the Tibetan population has aided their extreme environment adaptation [8].

Additionally, it is not only archaic genomes that are being explored, but also the computational reconstruction of the ancestral epigenetic state in hominins. The archaic DNA methylome can be indirectly estimated by quantifying the known degradation of methylated and unmethylated cytosines to thymine and uracil, respectively [9]. This has instigated a fascinating intersection of the fields of epigenetic, comparative and archaic genomics.

Epigenetic factors are principally the devices that each cell requires to regulate specialised functioning of its genome. They are the chemical marks and packaging of a genome that control the expression of the correct genes, at the appropriate time or under a particular condition. These include chemical modifications of DNA, post-translational modifications or variants of the histone proteins that DNA wraps around, and the complex interplay of certain non-coding RNA species [10]. This coordinated mechanism enables the cell to be propelled through development and into the specialised niche required for its synergistic role within the human body. The precise co-ordination of these mechanisms is therefore critical. Consequently, controlling the trajectory of global epigenetic modifiers is proposed to play a significant role in evolutionary change [11].

Epigenomic investigations may therefore lead to powerful functional insights, for example into the hidden complexity of the human brain [12]. A longstanding proposition is that epigenetic marks are a mechanism of molecular memory [13]. Decoding neuronal epigenomes and furthermore quantifying the plasticity of cellular states that enable complex higher functioning will help illuminate these biological processes. Therefore, primate comparison of local epigenetic factors as well as genetic changes in global modifying enzymes that mould the epigenome, will increase our understanding of human development and cognition [14,15]. However, it is not only neurological disorders but all aspects of our evolutionarily-acquired vulnerabilities that may be informed by these comparative studies, including metabolic, immune and musculoskeletal [16]. This review will survey recent insights into the function of the human epigenome from comparative primate studies. The importance of sequence variation in facilitating the epigenetic machinery will be a major focus.

2. Chromatin modifications

Chromatin comprises the dynamic packaging of DNA that also includes regulatory cues [17]. Evidence from yeast models indicate the proteins involved, histones, can themselves maintain their epigenetic state through somatic mitosis independently of associated DNA sequences [18,19]. Canonical histone modifications involve the modification of lysines (K) present in the tail of the histone 3 subunit (H3). These include signatures within latent promoter (H3K4me3) and enhancer regions (H3K4me1); an activity mark within both these regions (H3K27ac [20]); constitutive (H3K9me3) and facultative (H3K27me3) heterochromatin repressed regions [21]; as well as a combination of active and repressed marks indicating bivalent promoters [10]. Assaying a range of these chromatin marks has enabled combinatorial segmentation of the genome into demarcated tissue-specific functional regions [22]. The enhancer predictions [23], can be further functionally supported with associated expression of short enhancer-RNA (eRNA) [24].

The epigenetic machinery includes writers, erasers, readers and remodellers of the epigenome. Furthermore, it has been identified that there is a high mutational load in chromatin-modifying genes in developmental disorders [25]. These monogenic disorders strongly indicate the importance of chromatin dynamics and the associated phenotypes commonly including significant behavioural and intellectual disabilities [26]. It is also of note that these genes have also been observed as somatic hotspots in cancers [27]. This indicates their critical function in a time-independent fashion [28] and further underlines the importance of the accurate orchestration of the cellular epigenome.

3. DNA modifications

Epigenetic modifications of DNA include the most common epigenetic mark, methylation of the 5′ carbon of cytosine (5mC) [29]. In differentiated cells this occurs predominately in a CpG dinucleotide context, although non-CpG methylation does occur, particularly in the brain. The default state of CpGs throughout the genome is methylated [30], the notable exceptions being dense clusters of CpGs termed 'CpG islands' (CGIs) that are mostly unmethylated. The remainder low density and methylated CpGs are at an increased risk of loss over evolutionary time due to hypermutability and lack of recognition for repair [31]. Inversely, small increased GC clusters, including CpGs, can be generated as a by-product of recombination-associated Biased Gene Conversion (BGC) [31]. However, some lower density CpG regions may possess a reduced but not completely unmethylated DNA methylation state, termed Low Methylation Regions (LMRs). These regions co-locate with enhancer evidence including H3K4me1 and the enhancer-related co-activator p300 [32].

The CpG dinucleotide is proposed to act as a genome-wide signalling molecule in its own right, recruiting factors dependent upon its methylation state [33]. Furthermore distinct genetic motifs influence methylation state, particularly with CpG dense regions. These Methylation Determining Regions (MDR) include particular transcription factor binding site (TFBS) motifs, such as SP1, CTCF and members of the RFX family [34]. Other short sequences including AT motifs are suggested to also influence the epigenome [35].

The active removal of the DNA methylation mark occurs via an oxidative process catalysed by the TET enzymes [36]. This leads to the formation of hydroxymethylated cytosine (5hmC) with further oxidative steps deriving 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC), and then finally onto the unmodified cytosine base. The enzyme TDG may also be involved in these later changes. These modifications of DNA are recognised by particular readers [33]. Those that recognise 5mC include the methyl-CpG-binding protein (MECP2) and the methyl-CpG-binding-domain proteins (MBD1, MBD2 and MBD4), which enlist further molecules to create complexes that interact with chromatin remodellers [37].

The severe post-natal developmental decline that occurs in females with Rett syndrome is caused by mutation within the X-linked MECP2 gene. It clearly demonstrates the importance of this epigenetically-selective 5mC binding protein in the maturation of the central nervous system [38]. Furthermore, evidence in mice indicates it is required even in adulthood for correct neurological function [39] and additionally overexpression of MECP2 leads cynomolgus monkeys to display autistic behaviours [40]. Data has accrued that the MECP2 protein may also bind 5hmC in neuronal cells [41], as well as non-CpG methylated cytosines as the neuron matures [42,43], indicating the potential epigenetic complexity at work in the brain [44].

The plasticity of DNA methylation enhances the gene transcriptional repertoire of neurons and is proposed to influence the synaptic wiring implicated in memory [45]. In post-mitotic neurons widespread active DNA methylation removal occurs, with TET enzymes playing a prominent role in synaptic activity and downstream consequences on gene