Epigenomics

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/iepi20

Cell-Specific Gene Promoters are Marked by Broader Spans of H3K4me3 and are Associated with Robust Gene Expression Patterns

Duygu Ucar & Dashzeveg Bayarsaihan

To cite this article: Duygu Ucar & Dashzeveg Bayarsaihan (2015) Cell-Specific Gene Promoters are Marked by Broader Spans of H3K4me3 and are Associated with Robust Gene Expression Patterns, Epigenomics, 7:2, 129-131, DOI: 10.2217/epi.14.87

To link to this article: https://doi.org/10.2217/epi.14.87

Published online: 05 May 2015.

Editorial

For reprint orders, please contact: reprints@futuremedicine.com

Epigenomics

Cell-specific gene promoters are marked by broader spans of H3K4me3 and are associated with robust gene expression patterns

"...regions associated with buffer domains are also associated with a specific genomic landscape..."

Keywords: enhancers • epigenetic regulation • H3K4me3 • histone methylation

Genome-wide histone modification profiles contain precious information hidden in the spatial distribution of epigenetic marks. Several groups have observed that expanded regions of DNA sequences are marked by enhancer-specific histone modifications, such as H3K27ac, and coincide with cell-specific enhancers [1–4]. Similar observations have been made for DNA methylation profiles; extended regions of low DNA methylation (known as 'DNA methylation valleys or canyons') spanning at least 5 kilobase (kb) in length have been observed in human embryonic stem cells (hESCs) and hESC-derived cells, and have been associated with transcription factors and with genes known to regulate development [5]. Similarly DNA methylation canyons have been shown to mark genes with potential involvement in the regulation of hematopoiesis in hematopoietic stem cells [6].

More recently, Benayoun *et al.* identified a new subclass of trimethylation of histone H3 lysine 4 (H3K4me3) domains, namely 'buffer domains,' which preferentially mark genes associated with the underlying cell's identity and function [7]. The H3K4me3 modification marks the promoters of actively transcribed genes [8] and is shown to serve as a transcriptional on/off switch [9]. With the advancements in next-generation sequencing techniques, H3K4me3 mark has been extensively profiled in a large number of organisms and cell types. By taking advantage of the vast amount of publicly available H3K4me3 data sets, Benayoun and colleagues analyzed a large number of H3K4me3 ChIP-seq and ChIP-chip data sets (> 200 data sets). They noted that in each studied H3K4me3 data set, the length (i.e., breadth) distribution of H3K4me3 domains was characterized by a 'heavy right tail,' which refers to expanded genomic sites marked by H3K4me3 domains (spanning DNA regions from 4–5 kb up to 40–50 kb in length). Genes marked by these regions were observed to be critical for cell-specific functions in the cognate cell type, for example, pluripotency factor KLF4 is marked by a buffer domain in hESCs. The authors identified genes marked by these broad H3K4me3 domains in stem cells, in differentiated cells and in cancer cells derived from nine organisms, and they have generated a searchable database that will inform future efforts to identify the genes that are critical for studied cell types [7]. Further, they have shown that this subset of H3K4me3 domains can be effectively used to discriminate cell types and tissues with respect to their lineages and to prioritize genes with respect to their importance for the underlying cell type. For example, the authors have shown that known reprogramming factors are characterized by very long H3K4me3 marks in the cognate cell types, as in the case of *KLF4* in hESCs. In order to prove that buffer domains can be used for novel gene discovery, the authors generated H3K4me3 data sets in adult mouse neural progenitor cells (NPCs) and then ranked the genes in terms of their H3K4me3 domain length to identify buffer domains in NPCs. The

Duygu Ucar
Author for correspondence:
Department of Bioinformatics & Computational Biology, The Jackson Laboratory for Genomic Medicine, 10 Discovery Drive, Farmington, CT 06032, USA
duygu.ucar@jax.org

Dashzeveg Bayarsaihan
Department of Reconstructive Sciences, Institute for Systems Genomics & Center for Regenerative Medicine & Skeletal Development, University of Connecticut Health Center, 263 Farmington Avenue, Farmington, CT 06030, USA

Future Medicine part of fsg

authors then validated the efficacy of this epigenetic signature for novel gene discovery using a lentiviral-based RNA interference (shRNA) approach followed by cell proliferation and neurogenesis assays to quantify the ability of NPCs to proliferate and generate new neurons (neurogenesis). They identified several novel regulators with significant impact on neurogenesis, such as *SALL1*, *OTX1* and *BAHCC1*.

Benayoun *et al.* have additionally illustrated that these epigenetic biomarkers are not independent of the genomic and epigenomic landscape of the underlying loci. Using machine-learning models, they showed that regions associated with buffer domains are also associated with a specific genomic landscape – one that is conserved in human, mouse, worm and fly – that includes H3K79me2, TAF proteins, Pol2, cell-specific TF binding and transcriptional regulators SIN3A and CHD1. As the authors describe, this computationally derived signature represents a link between transcriptional elongation and broad H3K4me3 domains. To further unravel this association, the authors characterized the genes with broad H3K4me3 domains in terms of their Pol2 binding and gene expression patterns. Interestingly, they did not note any relationship between the level of gene expression and the breadth of H3K4me3 domains in neither single cell nor cell-population gene expression data sets. Intriguingly, they did note a relationship between the variance in gene expression levels (derived from replicates) and the breadth of the H3K4me3 domains. Based on these findings, the authors propose that a positive feedback mechanism is operating for key regulatory genes. In this mechanism, sustained elongation increases recruitment of H3K4me3-depositing complexes to the loci of cell-specific genes; this recruitment in turn promotes a broadening of H3K4me3 domains and recruits CHD1, thus facilitating the passage of elongating Pol2. The authors speculate that this feedback loop helps to ensure transcriptional consistency at the loci of cell-specific promoters.

To delve further into the process that leads to the formation and maintenance of these broad H3K4me3 domains, the authors knocked down *WDR5* with shRNA, which is an essential scaffolding subunit shared by all COMPASS/Trithorax-like complexes [10]. *WDR5* knockdown shortened the H3K4me3 domains in NPCs but did not significantly impact the total number of H3K4me3 domains present in these cells. Conversely, their analysis of publicly available *JARID1* knockdown data showed the expansion of H3K4me3 domain length. These analyses suggest existence of a complex chromatin writing/modifying mechanism that not only governs the deposit of an epigenetic mark but also maintains its length. The authors next coupled their *WDR5* knockdown experiments with corresponding RNA-seq experiments with replicates to measure changes in gene expression levels and variability after knockdown. These experiments indicated that changes in the length of H3K4me3 are correlated with the consistency of the expression of the marked gene. Further and more interestingly, by altering the length of a mark, they were also able to alter the consistency with which the underlying gene was expressed. Previous studies have described associations between levels of gene expression and several epigenetic marks (or combinations of marks); however, this is the first study to link a feature of an epigenetic mark (i.e., its length) to a specific output of gene expression (i.e., its consistent expression). The authors speculate that this conserved (and prevalent) H3K4me3 signature may ensure the consistent expression of key genes involved in cell identity and could protect against fluctuating environmental influences, a function that may be weakened by aging or disease.

Moving forward, the availability of assays for three-dimensional chromatin interactions (e.g., ChIA-PET) provides us with the tremendous opportunity to further understand the relationship between cell-specific enhancers and promoters in three-dimensional chromatin space. The mechanisms through which these buffer domains are selected and nucleated in diverse cells remain unexplored. The lincRNA–*WDR5* interface could represent a potential explanation as it has recently been shown that specific RNAs play integral roles in the WDR5–MLL complex in maintaining the active chromatin state and ensuring the fates of embryonic stem cells [11]. Another intriguing future investigation would be to vary the length of H3K4me3 domains, either via chemicals (e.g., similar to the use of JQ1 to inhibit super enhancers [12]) or via epigenome editing strategies [13], and then examine the regulatory impact of changes in the length of H3K4me3 mark across the genome. In closing, the buffer domain signature that has been recently discovered provides a simple and yet effective signature with the profound potential to inform the future discovery of key regulators in a variety of cell types and organisms and unravel an intriguing relationship between an epigenetic signature and the variability of gene expression [7].

Financial & competing interests disclosure

The authors have no relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript. This includes employment, consultancies, honoraria, stock ownership or options, expert testimony, grants or patents received or pending, or royalties.

No writing assistance was utilized in the production of this manuscript.

References

1 Whyte WA, Orlando DA, Hnisz D *et al.* Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* 153(2), 307–319 (2013).

2 Parker SC, Stitzel ML, Taylor DL *et al.* Chromatin stretch enhancer states drive cell-specific gene regulation and harbor human disease risk variants. *Proc. Natl Acad. Sci. USA* 110(44), 17921–17926 (2013).

3 Hnisz D, Abraham BJ, Lee TI *et al.* Super-enhancers in the control of cell identity and disease. *Cell* 155(4), 934–947 (2013).

4 Chapuy B, McKeown MR, Lin CY *et al.* Discovery and characterization of super-enhancer-associated dependencies in diffuse large B cell lymphoma. *Cancer Cell* 24(6), 777–790 (2013).

5 Xie W, Schultz MD, Lister R *et al.* Epigenomic analysis of multilineage differentiation of human embryonic stem cells. *Cell* 153(5), 1134–1148 (2013).

6 Jeong M, Sun D, Luo M *et al.* Large conserved domains of low DNA methylation maintained by Dnmt3a. *Nat. Genet.* 46(1), 17–23 (2014).

7 Benayoun BA, Pollina EA, Ucar D *et al.* H3K4me3 breadth is linked to cell identity and transcriptional consistency. *Cell* 158(3), 673–688 (2014).

8 Barski A, Cuddapah S, Cui K *et al.* High-resolution profiling of histone methylations in the human genome. *Cell* 129(4), 823–837 (2007).

9 Dong X, Greven MC, Kundaje A *et al.* Modeling gene expression using chromatin features in various cellular contexts. *Genome Biol.* 13(9), R53 (2012).

10 Migliori V, Mapelli M, Guccione E. On WD40 proteins: propelling our knowledge of transcriptional control? *Epigenetics* 7(8), 815–822 (2012).

11 Yang YW, Flynn RA, Chen Y *et al.* Essential role of lncRNA binding for WDR5 maintenance of active chromatin and embryonic stem cell pluripotency. *Elife* 3, e02046 (2014).

12 Loven J, Hoke HA, Lin CY *et al.* Selective inhibition of tumor oncogenes by disruption of super-enhancers. *Cell* 153(2), 320–334 (2013).

13 Mendenhall EM, Williamson KE, Reyon D *et al.* Locus-specific editing of histone modifications at endogenous enhancers. *Nat. Biotechnol.* 31(12), 1133–1136 (2013).