



Computational-based approaches in epigenetic research: Insights from computational tools, mathematical models, and machine learning methods

Roghayeh Ghorbani, Reza Shokri-Gharelo*

Department of Plant Breeding and Biotechnology, University of Tabriz, Tabriz, East Azarbaijan, Iran

ARTICLE INFO

Article history:

Received on: April 01, 2018

Accepted on: April 23, 2018

Available online: January 20, 2019

Key words:

Bioinformatics,
Computational epigenetics,
Data analysis,
DNA methylation,
Epigenomics

ABSTRACT

Understudying epigenetics underlying mechanisms is essential. Studies have indicated functional roles of epigenetic events, including DNA methylation and histone modifications, in human disease, stem cell growth and development, aging, response to environmental stresses, and species evolution. High-throughput sequencing techniques alongside routine experimental approaches have rapidly produced a bulk of data that the major part of them remained unprocessed. To analyze, interpret, and process of these data, availability of efficient computational methods is critical. Epigenetic data analysis is complex and difficult because these data contain a multi-layer set of information. The methods implemented for this purpose must be able to handle massive data of experimental works and process the epigenetic layers. Furthermore, the methods must be capable of integrating multiple modifications and their combination effects on chromatin conformational structure and consequently the expression network of genes. In this study, we briefly reviewed challenges in the way of the computational epigenetics, the latest reported methods, and significant biological results derived from applying computational-based methods on epigenetic data.

1. INTRODUCTION

The importance of epigenetics in biology and medicine is clear to every scientist who works in pertinent fields. Epigenetics refers to any mechanisms by which gene expression is altered without any changes in DNA sequence. Many studies have found an association of epigenetics with many diseases, stem cell functions, immunology, growth, and development, responding to stress, species conservation, evolution, aging, etc. [1-3]. These implications of epigenetics events have revealed essentiality of understanding epigenetics and epigenomics-epigenetic changes across the genome - especially its underlying molecular mechanisms.

DNA methylation is a well-known epigenetic process in which a methyl group (CH_3) is added on cytosine bases. The process particularly occurs at CpG islands (CGI) where cytosine bases are highly concentrated. The multiple DNA methyltransferase (DNMT) proteins mediate adding the methyl group at cytosine bases. Histone modification is another epigenetic process. The modifications of histones such

as methylation, acetylation, phosphorylation, sumoylation, and ubiquitination, alter the genome structural configuration, leading to gene expression changes at affected regions. The configurational changes provide/limit transcription factor accesses. Electrostatic interactions between histones and DNA are underlying powers to create structural changes [3-5]. The studies conducted during the past 20 years have indicated that the molecular mechanisms of causes and consequences of epigenetics are more complex and there is a need to investigate on many unknown prospects.

Some of the common experimental methods include chromatin immunoprecipitation (ChIP) (ChIP-on-chip combine ChIP) that is a method to detect differences between sample and control DNA. In this method, formaldehyde is first used to cross-linking DNA-bound proteins to the DNA. Then, histones are fragmented into about 500 base pairs, that each of these fragments with epigenetic modifications separated by antibodies. Finally, the nucleotides are released from the separated fragments and used to be hybridized with microarray data to find epigenetic modifications [6]. MeDIP (ChIP-on-chip that is called methyl-DNA immunoprecipitation [ChIP]) is a varied method of ChIP in which antibodies are used against methylated cytosine. This method is widely used to find DNA methylation [7]. ChIP-seq uses high-throughput DNA sequencing instead of hybridization process and is more accurate and cheaper than its primary forms, described earlier [8]. More detailed information about mentioned experimental methods could be found in a review written by Schones and Zhao [9].

*Corresponding Author

Reza Shokri-Gharelo,

Department of Plant Breeding and Biotechnology,

University of Tabriz, 29 Bahman Blvd.,

Tabriz 5166616471, East Azarbaijan, Iran.

Tel.: +98-935-601-8541.

Email: shokrireza67@gmail.com

Bisulfite sequencing is a high-resolution method that is able to detect DNA methylation patterns. In spite of high resolution and efficiency, this method is expensive [10]. All of these methods integrated with several protocols have widely been used. The majority of protocols to assess epigenetic events are based on some of the following basics: (i) Techniques to inhibit DNA methylation and assess DNA methylation activity, (ii) ChIP-based protocols, (iii) *in vivo* RNA-Protein interaction assessment, and (iv) knockdown of histone deacetylases [11]. Besides experimental works, bioinformatics tools and computational biology have made remarkable advances in elucidating epigenetic events. Various bioinformatic tools and computational methods have been developed for managing, handling, and analyzing different epigenetic data, and many researches have been conducted in this context [Figure 1]. The importance of computational-based methods in the epigenetic fields has essentially been revealed after emerging high-throughput sequencing techniques that have accumulated the bulk of unannotated data. Indeed, computational methods are used for two main purposes: First, for interpreting data resulted from experimental works, and second, for analyzing the genome to find specific sequences that are responsible for epigenetic modifications such as CGIs. Increasing large-scale dataset has enriched analysis related to epigenetics, but also it has made it more complex and further outline crucial roles of computational tools in this regard. Here, we briefly reviewed the computational epigenetics and the latest bioinformatic methods developed to study epigenetic causes and consequences.

2. THE CHALLENGES AND OPPORTUNITIES OF COMPUTATIONAL EPIGENETICS

Computational epigenetics is constituted of different sciences including computer science, statistics, physics, and computational biology. The aim of computational epigenetics is to design and develop computational-based methods and programs to analyze data resulted from experimental works on epigenetics [12]. Data of epigenetic research encompass multiple layers of regulatory mechanisms and clues that must mainly be extracted from high-throughput sequencing techniques [13]. The methods must address available issues related to (1) the experimental methods such as background read problem in

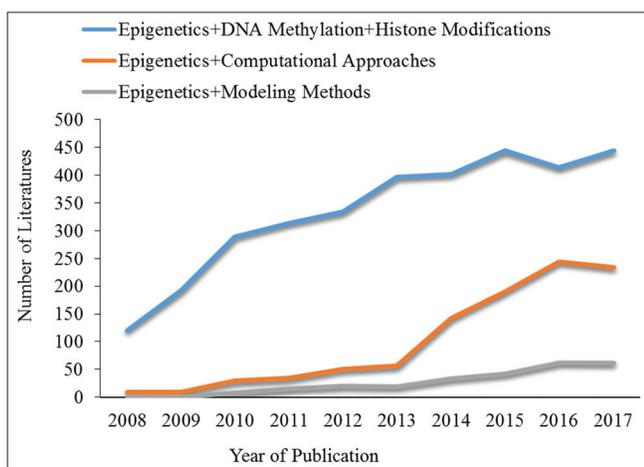


Figure 1: Number of literature related to epigenetics and computational approaches published between the years 2008 and 2017. The keywords used for the search were epigenetics, DNA methylation, histone modifications, computational approaches, and modeling methods. A number of hits for epigenetics+machian learning search were 65 literatures. The search was performed at PubMed (<https://www.ncbi.nlm.nih.gov/pubmed>)

ChIP-seq [14], (2) analytic approaches [15], (3) methods to integrate interplay of other compounds, such as microRNAs, with epigenetic regulation [16] and (4) the biased data resulted from the experimental mythologies such as profiling DNA methylation by MBD-seq [17]. In addition, there is more space for computational methodologies to open insight on protein-protein interactions, to decrease the cost of epigenome mapping, to theoretically modeling of epigenetic mechanisms, and to improve statistical genome browsers [18].

3. DATA SOURCES

Databases are one of the main places that information related to epigenetics could be accessed. As tabulated in Table 1, there are many databases for epigenetics. MethDB (<http://www.methdb.de>) contains information about DNA methylation and methylation patterns in many species. PubMed (www.pubmed.org/) contains a text-type file of more than thousands scientific articles related to methylated and other types of epigenetic modifications. REBASE (<http://rebase.neb.com/rebase/rebase.html>) is a database that connected to GenBank database. Genes responsible to encode DNMT have deposited in this database. Epigenetic data related to human chromatin and disease could reach in MeInfoText (<http://mit.lifescience.ntu.edu.tw/>), MethPrimerDB (<http://medgen.ugent.be/methprimerdb/>), The Krembil Family Epigenetics Laboratory (<http://www.epigenomics.ca>), and MethyLogiX DNA methylation database (<http://www.methylogix.com/genetics/database.shtml.htm>). Of other useful databases, we could mention to the Histone Database, ChromDB, CREMOFAC, the National Human Genome Research Institute's Histone Database, the National Center for Biotechnology Information's Gene Expression Omnibus, the Gene Normal Tissue Expression database, DNA Data Bank of Japan, and the Blood express database [12].

4. COMPUTATIONAL ANALYSIS OF EPIGENETIC DATA

There is a variety of computational-based approaches to analyze, modulate, and predict epigenetic modifications in given sequences. In the case of detecting DNA methylation, efforts have mainly devoted to discover of methylated CGIs and allele-specific cytosine methylation. The CG dinucleotides are mostly scarce throughout the genome, especially in vertebrates [19] and mainly clustered in the regions called CGIs. These rich regions with CG dinucleotides, CGIs, are interestingly located at the promoter of coding and non-coding genes, making them very attractive for researchers [19,20]. Because altering DNA methylation patterns of CGIs play essential roles in controlling the gene expression and silencing in various biological processes, such as X-chromosome inactivation, imprinting, silencing of intragenomic parasites [21,22], and especially in the epigenetic causes of cancer [21]. Due to CGIs essential implications in mentioned processes, multiple algorithms (either specific species or general purpose) have been developed to identify CGIs in the genomes. In this context, an algorithm to study CGIs and G+C content in the genome of vertebrates was first used by Gardiner-Garden and Frommer [23]. Subsequently, many other methods based on different algorithms had been developed. Of these methods, artificial neural networks (ANN) and support vector machines (SVM) have broadly been used to analyze DNA methylation. Marchevsky *et al.* [23] trained ANN with molecular data to classify lung cancer cells based on DNA methylation marker. They provided evidence that ANN could be used as a powerful approach for detecting DNA methylation. Das *et al.* [24] indicated that SVM could predict methylation status of CpG regions with an accuracy of 86%. They used this method to depict methylation patterns of all 22 human autosome chromosomes. The methods such as hidden Markov models (HMM),

Table 1: Some of the important databases that provide information about epigenetic events in various organisms

Database name	Description	Access address
MethDB	DNA methylation	http://www.methdb.de
PubMed	Text-type file of scientific articles related to methylated and other types of epigenetic modifications	www.pubmeth.org/
REBASE	Genes responsible to encode DNMT	http://rebase.neb.com/rebase/rebase.html/
MeInfoText	Epigenetic data related to human chromatin and disease	http://mit.lifescience.ntu.edu.tw/
MethPrimerDB	PCR primer database for DNA methylation analysis	http://medgen.ugent.be/methprimerdb/
CAMH	Epigenetic modifications of DNA and chromosomal proteins histones related to human diseases	http://www.epigenomics.ca
MethyLogiX	DNA methylation	http://www.methylogix.com/genetics/database.shtml.htm
The histone database	Full-length sequences and structures of histones and nonhistone proteins	https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1800941/
ChromDB	Chromatin-associated proteins	http://www.chromdb.org/
CREMOFAC	A database of chromatin remodeling factors	http://www.jncasr.ac.in/cremofac/
NHGRI's histone database	Histone proteins and their sequence variants	https://www.ncbi.nlm.nih.gov/research/HistoneDB2.0/index.fcgi/browse/
NCBI's GEO	Microarray, next-generation sequencing, and other forms of high-throughput functional genomic data	https://www.ncbi.nlm.nih.gov/geo/
GeneNote	An expression profile for genes based on high-density DNA microarrays	http://genecards.weizmann.ac.il/genenote/
DDBJ	A center for sharing and analysis services for data	https://www.ddbj.nig.ac.jp/index-e.html
The BloodExpress database	A database of gene expression in mouse hematopoiesis	http://hscl.cimr.cam.ac.uk/bloodexpress/

CAMH: Krembil Family Epigenetics Laboratory, NHGRI's Histone Database, the National Human Genome Research Institute (NHGRI)'s Histone Database, NCBI's GEO: The National Center for Biotechnology Information (NCBI)'s Gene Expression Omnibus (GEO), GeneNote: The Gene Normal Tissue Expression (GeneNote) database, DDBJ: DNA Data Bank of Japan

logistic regression, K-nearest neighbors, and decision trees have also been used for this purpose [25,26], for example, Barazandeh *et al.* [19] disclosed significant correlations between CGI density and genomic features such as chromosome size, GC content, ObsCpG/ExpCpG, gene density, and recombination rate in cattle. However, these methods were suffering from several disadvantages. First, these methods lacked systematic selection methods for a length threshold [27]. Second, they were unable to detect weak CGIs [28]. Third, they were sequence-based to identify CGIs and failed to distinguish between genuine CGIs and CpG-rich regions [29]. To overcome these drawbacks, Bock *et al.* [30] suggested epigenome prediction method and used integrates DNA methylation, Polymerase II preinitiation complex binding, histone H3K4 di- and trimethylation, histone H3K9/14 acetylation, DNase I hypersensitivity, and SP1 binding as criteria to map CGIs. Their method could distinguish between weak and stronger CGIs and use features of genomic DNA sequences and epigenome.

In respect to modifications of CGIs, the methylation is not only modification. Studies on the mammalian genome have demonstrated that in addition to methylation, there are other forms of modification including hydroxymethylation, formylation, and carboxylation [31,32]. The specific roles of these type modifications are still little-known, but it has been hypothesized that these modifications might be intermediate steps during methylation and demethylation processes or even they may have own implication in diseases [33,34]. Furthermore, studies on three high-resolution structures of chromatin have revealed the effect of two methylations, two hydroxymethylation, and five formylations on DNA dodecamers, while methylation and hydroxymethylation alone have not any effects on the geometry of DNA [35,36]. These imply this fact that for fully understanding of CGI modification effects on chromatin's structure; consequently, altering expression of genes it is critical that analyzing methods must include all of the modification

possibilities. One of the computational methods that meet these criteria is reported by Krawczyk *et al.* [37] in which they extended Natural Move Monte Carlo to simulate the conformation changes of chromatin as consequence of epigenetic modifications.

In the case of prediction, modeling, and analysis of histone modifications, some methods have reported, such as simplified stochastic model [38], genome-wide chromatin analysis [39], and genome-wide mapping [40]. From the machine-learning methods have been used to detect histone modifications (acetylation, methylation, and phosphorylation) we can mention to HMM approach [41], using chromatin signatures [42], model based on the prediction of pH-dependent aqueous solubility [43], and HMM based on the domain-level behavior [44]. Benveniste *et al.* [45] recently showed that histone modification prediction could achieve from knowledge of transcription factor binding at both promoter and distal regulatory elements. Furthermore, the methods such as QSAR analysis, homology modeling, and molecular docking methods have used for detecting histone modifications [46,47]. These tools have well been used for deciphering epigenetic effects on various biological processes. Furthermore, using the interaction between epigenetics, genetics, and environment can improve estimation of breeding values and reduce their biases [48].

The histone code hypothesis [49] articulates that the roles of histone post-translational modifications (PTM) are well described when the combinations and sequences of histone PTMs are accounted. Based on this hypothesis, several computational methods were developed for identification of histone modifications. ChromaSig and ChromHMM are two computational methods that have been developed for histone modifications [50,51]. These methods are based on multivariate HMM and are able to show histone modifications and chromatin

statuses. Given that only subsets of the histone PTM combinations take place in nature, the later approaches were developed based on partial correlations and maximum entropy modeling. These methods have been used for identification of pairwise and high-order interactions between chromatin factors [52].

5. COMPUTATIONAL EPIGENETIC RESEARCH

Computational methods have remarkably helped to explore molecular mechanisms of epigenetics and its association with biological processes. Cancer is one of the fields that computational epigenetics has widely been used. It has been indicated that DNA methylation patterns in cancer are variable and tumor type specific. To elucidate DNA methylation pattern in cancer, several computational approaches have been used, such as using regression models to analyze DNA methylation profile [53], using SVM to analyze DNA methylation in tumor class [54], and using Manhattan distance and average linkage algorithms for CGI pattern analysis of human colorectal tumors [55].

Stem cell is another field that computational epigenetics has widely used. Recent studies have revealed unique epigenetic profiles of embryonic stem cells, as reviewed by Spivakov and Fisher [56]. Walker *et al.* [57] formulated novel networks that indicate gene response of key developmental regulators in embryonic stem cell and could predict the outcomes of genetic manipulation in this network. They used temporal expression microarray analyses and known genome-wide transcription factor to construct the networks. In another study, Ringrose *et al.* [58] used computational-based method to successfully identify 167 candidate polycomb/trithorax response elements. These elements are involved in the development and cell proliferation.

Neurodegenerative and autoimmune diseases are two another important diseases that have been studied using a computational-based method to find epigenetic factors responsible for the diseases. Khanam Irin *et al.* [59] in this study proposed a computational method able to explain the functional consequences of epigenetic modification. The method, called biological expression language, is capable of integrating literature-derived information into network model. Moreover, it is possible to apply reverse causal reasoning algorithms, which support the identification of mechanistic hypothesis from related network model.

6. CONCLUDING REMARKS AND OUTLOOK

Many high-throughput sequencing technologies open a new era for epigenetic research. To handle millions of these data, many computational tools have been developed. There are, however, issues that the computational tools have to address. Computational methods must be applied to the integrative analysis of epigenetic layers. These methods could remarkably improve our knowledge of complex regulatory processes and interconnections by which epigenetics works. It has been observed that even small networks with few components in epigenetic events tend to behave in complex and unexpected manners. Therefore, there is a need to build up systematic and focused modular approaches to elucidate fundamental understanding of epigenetics. Some methods developed for analyzing literature-derived data are not efficient in showing epigenetic modifications at gene level so extending these methods should be considered.

It is expected that the computational methodologies will shift for being able to interpret data so that can be used in quantifying the disease risks and driving therapeutics. They should draw meaningful inferences of epigenetic modifications in diseases, and develop

novel approaches for new powerful epigenome-editing and high-throughput experimental methodologies. The new computational methodologies should integrate the combination of the computational methods, especially machine learning approaches. Considering these strategies in developing new computational methods could extend our understanding of epigenetic mechanisms.

7. REFERENCES

- Handel AE, Ebers GC, Ramagopalan SV. Epigenetics: Molecular mechanisms and implications for disease. *Trends Mol Med* 2010;16:7-16.
- Chinnusamy V, Zhu JK. Epigenetic regulation of stress responses in plants. *Curr Opin Plant Biol* 2009;12:133-9.
- Weinhold B. Epigenetics: The science of change. *Environ Health Perspect* 2006;114:A160-7.
- Lima R, Hayashi D, Lima K, Gomes N, Ribeiro M. The role of epigenetics in the etiology of obesity: A review. *J Clin Epigenet* 2017;3:41.
- Allis CD, Jenuwein T. The molecular hallmarks of epigenetic control. *Nat Rev Genet* 2016;17:487-500.
- Buck MJ, Lieb JD. ChIP-chip: Considerations for the design, analysis, and application of genome-wide chromatin immunoprecipitation experiments. *Genomics* 2004;83:349-60.
- Jacinto FV, Ballestar E, Esteller M. Methyl-DNA immunoprecipitation (MeDIP): Hunting down the DNA methylome. *Biotechniques* 2008;44:35, 37, 39 passim.
- Park PJ. ChIP-seq: Advantages and challenges of a maturing technology. *Nat Rev Gen* 2009;10:669-80.
- Schones DE, Zhao K. Genome-wide approaches to studying chromatin modifications. *Nat Rev Genet* 2008;9:179-91.
- Hajkova P, el-Maarri O, Engemann S, Oswald J, Olek A, Walter J, *et al.* DNA-methylation analysis by the bisulfite-assisted genomic sequencing method. *Methods Mol Biol* 2002;200:143-54.
- Tollefsbol TO. *Epigenetics Protocols*. Vol. 287. Springer Science and Business Media; 2004.
- Lim SJ, Tan TW, Tong JC. Computational epigenetics: The new scientific paradigm. *Bioinformatics* 2010;4:331-7.
- Robinson MD, Pelizzola M. Computational epigenomics: Challenges and opportunities. *Front Genet* 2015;6:88.
- Flensburg C, Kinkel SA, Keniry A, Blewitt ME, Oshlack A. A comparison of control samples for chIP-seq of histone modifications. *Front Genet* 2014;5:329.
- Robinson MD, Kahraman A, Law CW, Lindsay H, Nowicka M, Weber LM, *et al.* Statistical methods for detecting differentially methylated loci and regions. *Front Genet* 2014;5:324.
- Osella M, Riba A, Testori A, Corà D, Caselle M. Interplay of microRNA and epigenetic regulation in the human regulatory network. *Front Genet* 2014;5:345.
- Mensaert K, Van Criekinge W, Thas O, Schuurings E, Steenbergen RD, Wisman GB, *et al.* Mining for viral fragments in methylation enriched sequencing data. *Front Genet* 2015;6:16.
- Bock C, Lengauer T. Computational epigenetics. *Bioinformatics* 2008;24:1-0.
- Barazandeh A, Mohammadabadi M, Ghaderi-Zefrehei M, Nezamabadi-Pour H. Genome-wide analysis of CpG islands in some livestock genomes and their relationship with genomic features. *Czech J Anim Sci* 2016;61:487-95.
- Hackenberg M, Barturen G, Carpena P, Luque-Escamilla PL, Previti C, Oliver JL, *et al.* Prediction of cpG-island function: CpG clustering vs. Sliding-window methods. *BMC Genomics* 2010;11:327.
- Barazandeh A, Mohammadabadi M, Ghaderi-Zefrehei M, Nezamabadipour H. Predicting CpG islands and their relationship with genomic feature in cattle by hidden markov model algorithm.

- Iran J Appl Anim Sci 2016;6:571-9.
22. Su J, Zhang Y, Lv J, Liu H, Tang X, Wang F, *et al.* CpG_MI: A novel approach for identifying functional cpG islands in mammalian genomes. *Nucleic Acids Res* 2010;38:e6.
 23. Marchevsky AM, Tsou JA, Laird-Offringa IA. Classification of individual lung cancer cell lines based on DNA methylation markers: Use of linear discriminant analysis and artificial neural networks. *J Mol Diagn* 2004;6:28-36.
 24. Das R, Dimitrova N, Xuan Z, Rollins RA, Haghghi F, Edwards JR, *et al.* Computational prediction of methylation status in human genomic sequences. *Proc Natl Acad Sci U S A* 2006;103:10713-6.
 25. Bhasin M, Zhang H, Reinherz EL, Reche PA. Prediction of methylated cpGs in DNA sequences using a support vector machine. *FEBS Lett* 2005;579:4302-8.
 26. Chen H, Xue Y, Huang N, Yao X, Sun Z. MeMo: A web tool for prediction of protein methylation modifications. *Nucleic Acids Res* 2006;34:W249-53.
 27. Takai D, Jones PA. Comprehensive analysis of cpG islands in human chromosomes 21 and 22. *Proc Natl Acad Sci U S A* 2002;99:3740-5.
 28. Weber M, Hellmann I, Stadler MB, Ramos L, Pääbo S, Rebhan M, *et al.* Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. *Nat Genet* 2007;39:457-66.
 29. Yamada Y, Watanabe H, Miura F, Soejima H, Uchiyama M, Iwasaka T, *et al.* A comprehensive analysis of allelic methylation status of cpG islands on human chromosome 21q. *Genome Res* 2004;14:247-66.
 30. Bock C, Walter J, Paulsen M, Lengauer T. CpG island mapping by epigenome prediction. *PLoS Comput Biol* 2007;3:e110.
 31. Bhutani N, Burns DM, Blau HM. DNA demethylation dynamics. *Cell* 2011;146:866-72.
 32. Wu H, Zhang Y. Charting oxidized methylcytosines at base resolution. *Nat Struct Mol Biol* 2015;22:656-61.
 33. Kroeze LI, van der Reijden BA, Jansen JH 5-hydroxymethylcytosine: An epigenetic mark frequently deregulated in cancer. *Biochim Biophys Acta* 2015;1855:144-54.
 34. Guo JU, Su Y, Zhong C, Ming GL, Song H. Emerging roles of TET proteins and 5-hydroxymethylcytosines in active DNA demethylation and beyond. *Cell Cycle* 2011;10:2662-8.
 35. Drew HR, Wing RM, Takano T, Broka C, Tanaka S, Itakura K, *et al.* Structure of a B-DNA dodecamer: Conformation and dynamics. *Proc Natl Acad Sci U S A* 1981;78:2179-83.
 36. Lercher L, McDonough MA, El-Sagheer AH, Thalhhammer A, Kriacionis S, Brown T, *et al.* Structural insights into how 5-hydroxymethylation influences transcription factor binding. *Chem Commun (Camb)* 2014;50:1794-6.
 37. Krawczyk K, Demharter S, Knapp B, Deane CM, Minary P. *In silico* structural modeling of multiple epigenetic marks on DNA. *Bioinformatics* 2018;34:41-8.
 38. Dodd IB, Micheelsen MA, Sneppen K, Thon G. Theoretical analysis of epigenetic cell memory by nucleosome modification. *Cell* 2007;129:813-22.
 39. Schübeler D, MacAlpine DM, Scalzo D, Wirbelauer C, Kooperberg C, van Leeuwen F, *et al.* The histone modification pattern of active genes revealed through genome-wide chromatin analysis of a higher eukaryote. *Genes Dev* 2004;18:1263-71.
 40. Roh TY, Cuddapah S, Zhao K. Active chromatin domains are defined by acetylation islands revealed by genome-wide mapping. *Genes Dev* 2005;19:542-52.
 41. Xu H, Wei CL, Lin F, Sung WK. An HMM approach to genome-wide identification of differential histone modification sites from chIP-seq data. *Bioinformatics* 2008;24:2344-9.
 42. Won KJ, Chepelev I, Ren B, Wang W. Prediction of regulatory elements in mammalian genomes using chromatin signatures. *BMC Bioinformatics* 2008;9:547.
 43. Kouskoumvekaki I, Hansen NT, Björkling F, Vadlamudi SM, Jónsdóttir SO. Prediction of pH-dependent aqueous solubility of histone deacetylase (HDAC) inhibitors. *SAR QSAR Environ Res* 2008;19:167-77.
 44. Thurman RE, Day N, Noble WS, Stamatoiyannopoulos JA. Identification of higher-order functional domains in the human ENCODE regions. *Genome Res* 2007;17:917-27.
 45. Benveniste D, Sonntag HJ, Sanguinetti G, Sproul D. Transcription factor binding predicts histone modifications in human cell lines. *Proc Natl Acad Sci U S A* 2014;111:13367-72.
 46. Juvele DC, Kulkarni VV, Deokar HS, Wagh NK, Padhye SB, Kulkarni VM, *et al.* 3D-QSAR of histone deacetylase inhibitors: Hydroxamate analogues. *Org Biomol Chem* 2006;4:2858-68.
 47. Lin YC, Lin JH, Chou CW, Chang YF, Yeh SH, Chen CC, *et al.* Statins increase p21 through inhibition of histone deacetylase activity and release of promoter-associated HDAC1/2. *Cancer Res* 2008;68:2375-83.
 48. Roudbar MA, Mohammadabadi M, Salmani V. Epigenetics: A new challenge in animal breeding. *Gen Third Millennium* 2014;12:3900-14.
 49. Strahl BD, Allis CD. The language of covalent histone modifications. *Nature* 2000;403:41-5.
 50. Hon G, Ren B, Wang W. ChromaSig: A probabilistic approach to finding common chromatin signatures in the human genome. *PLoS Comput Biol* 2008;4:e1000201.
 51. Ernst J, Kellis M. ChromHMM: Automating chromatin-state discovery and characterization. *Nat Methods* 2012;9:215-6.
 52. Zhou J, Troyanskaya OG. Global quantitative modeling of chromatin factor interactions. *PLoS Comput Biol* 2014;10:e1003525.
 53. Bock C, Walter J, Paulsen M, Lengauer T. Inter-individual variation of DNA methylation and its implications for large-scale epigenome mapping. *Nucleic Acids Res* 2008;36:e55.
 54. Adorján P, Distler J, Lipscher E, Model F, Müller J, Pelet C, *et al.* Tumour class prediction and discovery by microarray-based DNA methylation analysis. *Nucleic Acids Res* 2002;30:e21.
 55. Weisenberger DJ, Siegmund KD, Campan M, Young J, Long TI, Faasse MA, *et al.* CpG island methylator phenotype underlies sporadic microsatellite instability and is tightly associated with BRAF mutation in colorectal cancer. *Nat Genet* 2006;38:787-93.
 56. Spivakov M, Fisher AG. Epigenetic signatures of stem-cell identity. *Nat Rev Genet* 2007;8:263-71.
 57. Walker E, Ohishi M, Davey RE, Zhang W, Cassar PA, Tanaka TS, *et al.* Prediction and testing of novel transcriptional networks regulating embryonic stem cell self-renewal and commitment. *Cell Stem Cell* 2007;1:71-86.
 58. Ringrose L, Rehmsmeier M, Dura JM, Paro R. Genome-wide prediction of polycomb/Trithorax response elements in drosophila melanogaster. *Dev Cell* 2003;5:759-71.
 59. Khanam Irin A, Kodamullil AT, Gündel M, Hofmann-Apitius M. Computational modelling approaches on epigenetic factors in neurodegenerative and autoimmune diseases and their mechanistic analysis. *J Immunol Res* 2015;2015:737168.

How to cite this article:

Ghorbani R, Shokri-Gharelo R. Computational-based approaches in epigenetic research: Insights from computational tools, mathematical models, and machine learning methods. *J App Biol Biotech.* 2019;7(01):79-83. DOI: 10.7324/JABB.2019.70114