Gaussian derivative wavelets identify dynamic changes in histone modification

Nha Nguyen
Department of Genetics, Institute for Diabetes, Obesity and Metabolism, School of Medicine, University of Pennsylvania, Philadelphia
Email: nhnguyen@mail.med.upenn.edu

Kyoung-Jae Won
Department of Genetics, Institute for Diabetes, Obesity and Metabolism, School of Medicine, University of Pennsylvania, Philadelphia
Email: wonk@mail.med.upenn.edu

Abstract— **Epigenetic landscapes reveal how cells regulate genes in a cell-type or condition specific manner. Genome-wide surveys using histone modification showed cell-type specific regulatory regions. A number of computational methods were designed to identify cell-type specific regulatory regions using epigenome data. Most of them were designed to identify the enrichment of histone modification or their changes. However, they did not consider the shape of epigenetic signals, which represents the condition for protein binding at gene regulatory regions.**

We present a computational method to detect epigenetic changes using the shape of the signals for histone modification. Employing a Gaussian Derivative Wavelet (CGDWavelet) approach, the proposed method models a nucleosome with a Gaussian and detects the peak and the edges of the Gaussian. Using the detected parameters across two samples, CGDWavelet classifies epigenetic changes. We applied CGDWavelet to the histone modification data from mouse embryonic stem cells (mESCs) and neural progenitor cells (mNPCs) and identified four groups of epigenetic changes. Associating each group with gene expression, we found that gene expression is affected by chromatin structure as well as the intensity of histone modification. We found that Smad1, Sox2 and Nanog but not Oct4 bind to the epigenetically variable regions for H3K4me3. Software is available at http://wonk.med.upenn.edu/CGDWavelet

Keywords-component; wavelet, epigenome, histone modification, nucleosome

I. INTRODUCTION

Gene regulation is orchestrated by many factors including transcription factors (TFs), their co-factors, RNA polymerase as well as epigenetic status. The epigenetic landscapes, represented by modifications to histones, DNA methylation and other proteins that package the genome, regulate the function of cells by activating or repressing gene activity [1, 2]. Epigenetic changes represent the changes in the environment for gene regulation. Epigenetic changes during cell differentiation reflect the commitment to a lineage, leading to cell-type specific gene regulation [4, 5]. Environmental changes also affect the epigenetic landscapes. Treatment of androgen receptor agonist to the prostate cancer cells changed the positions of nucleosomes as well as the associated histone modification patterns [6].

Large portion of genomic regions are enriched for various epigenetic marks. Promoters, regulatory regions around gene transcription start sites (TSSs) are associated with active histone mark such as tri-methylation of Lys4 of H3 (H3K4me3) and histone acetylation when it is active. Enhancers, distal regulatory region for gene regulation, are enriched for mono-, di-, and tri-methylation of Lys4 of H3 (H3K4me1/2) and histone acetylation [7, 8]. Repressed genes are enriched for repressive histone modification marks such as H3K27me3, H3K9me3 and H3K20me3. Histone modifications were used to predict regulatory regions and transcription factor binding sites (TFBSs) [9-12]. At TFBSs, bimodal histone modification patterns were observed, showing a nucleosome free region at the center. DNaseI hypersensitive sites, therefore, were successfully used to predict TFBSs. He et al. successfully identified dynamic nucleosome changes using H3K4me2 and found FoxA1 binding sites [6]. At promoters, a dynamic equilibrium switch mechanism suggests the competition between nucleosomes and TFs [13-15]. These lines of evidence suggest the importance of epigenetic changes for condition-dependent gene regulation. Also, it is required to develop computational algorithms to identify epigenetic changes to understand condition-specific gene regulation.

In this paper, we suggest a method, called CGDWavelet, one extension version of our method in [5] , which uses Gaussian derivative wavelets to detect epigenetic changes. CGDWavelet identifies the shapes of histone modification and evaluates their changes across two samples. To focus on the epigenetic changes at gene regulatory regions, CGDWavelet uses active histone marks such as H3K4me1/2/3 or any histone acetylation and identifies bimodal peaks and their changes. Previously, we developed AWNFR which detects nucleosome position based on histone modification data using wavelet transform [5]. Inherited from AWNFR, CGDWavelet uses wavelet to identify the parameters of a Gaussian function that models a nucleosome. Compared with AWNFR which uses mixture of Gaussian to model a nucleosome [5], CGDWavelet uses a simple Gaussian to improve the speed of implementation. Algorithmically, CGDWavelet is equipped with 1-dimensional (1D) scalogram to detect bimodal peaks more efficiently and provide more parameters for subsequent classification of epigenetic changes.

A number of algorithms were developed to detect nucleosome positions or nucleosome free regions (NFRs) using histone modification [9, 16-18]. Compared with them, CGDWavelet was designed to detect dynamic changes of

nucleosome without any post-processing. Compared with [6], which detects dynamic nucleosome positions using signal intensities, CGDWavelet employed wavelet transform for accurate prediction of nucleosome positions. ChIPDiff, DHMS and diffReps were developed to detect the region with epigenetic changes [19-21]. While they focused on the differential level of histone modification signals at certain regions, CGDWavelet is focused on the changes of the epigenetic patterns.

Using CGDWavelet, we investigated the epigenetic changes from mouse embryonic stem cell (mESCs) to mouse neural progenital cells (mNPCs). Based on the obtained Gaussian parameters, CGDWavelet classifies epigenetic changes into 4 groups: 1) loss of bimodal peak, 2) only the histone modification levels are changed (or no substantial changes) 3) slight closure of the chromatin structure and 4) full closure of the chromatin structure. We found that gene expression was closely related with opening of chromatin structure. We also studied TFs associated with each group of epigenetic changes.

II. METHOLODY

CGDWavelet uses a Gaussian derivative wavelet method to detect nucleosome positions marked by histone modifications. CGDWavelet decomposes the histone modification data and converts them into the wavelet domain. After assuming a Gaussian for a nucleosome, CGDWavelet calculates the edges and the peaks of the Gaussian using zero-crossing lines and 1D wavelet [3, 22]. The obtained parameters of the Gaussians are used to detect nucleosomes and their dynamic changes.

A. One-directional wavelet scholargram

The wavelet transform of histone modification signal can be rewritten as a multi-scale differential operator

$$W_n f(u,s) = s^n \frac{d^n}{du^n}(f_i \star \bar{\theta}_s(t))(u), \quad (1)$$

where, the Gaussian wavelet is,

$$\bar{\theta}_s(t) = \frac{1}{\sqrt{s}} e^{-\frac{t^2}{s^2}} \quad (2)$$

Convoluting $f_i(t)$ and $\bar{\theta}_s(t)$, we get

$$(f_i \star \bar{\theta}_s)(u) = K_1 e^{-K_2(u-\mu_i)^2}, \quad (3)$$

where $K_1 = A\sqrt{\frac{2\pi\sigma_i^2 s}{2\sigma_i^2+s^2}}$ and $K_2 = \frac{1}{2\sigma_i^2+s^2}$, respectively.

The first derivative of (3) is

$$W_1 f(u,s) = -2sK_1K_2(u-\mu_i)e^{-K_2(u-\mu_i)^2}. \quad (4)$$

$W_1 f(u,s) = 0$ when $u_0 = \mu_i$ and $u_0(s+1) - u_0(s) = 0$ for any scale *s*. The zero points, u_0, across scales result in zero-crossing lines whose positions is corresponding to the peaks of histone modification signals. If the zero-crossing line is a continuous line with length N, the positions of the peaks are

$$\mu_i = \frac{1}{N}\sum_{s=1}^{N} u_0(s).$$

$W_1 f(u,s)$ has its maximum at $\mu_0 = \mu_i \pm \sqrt{\sigma_i^2 + \frac{s^2}{2}}$.

The scalogram in 2D is

$$WS(u,s) = 100 \times \frac{(\frac{W_1 f(u,s)}{\sqrt{s}})^2}{\sum_{i=1}^{N}(\frac{W_1 f(u,s)}{\sqrt{s}})^2}. \quad (5)$$

However, detecting breakpoints with 2D scalogram is computationally exhaustive. To solve this, we changed the 2D scalogram into 1D scalogram. For this, ridge lines [23] were identified by linking the local maxima of 2D scalogram at each scale level. We denote L_R and $\mathcal{U}(u)$ as linking line length and a vector for linked maxima at for u. Also in this step, the ridge line with a length smaller than a certain threshold is set to zero.

$$\mathcal{U} = \begin{cases} 0, \ if\ L_R < threshold \\ u_1\, u_2 \ldots u_{s_{max}}, \ otherwise \end{cases} \quad (6)$$

Then, 1D scalogram can be described as

$$WS_{1D}(\bar{u}) = \begin{cases} 0, \ if\ \mathcal{U} = 0, \\ \sum_{u\in\mathcal{U}} WS(u,s), \ otherwise, \end{cases} \quad (7)$$

where

$$\bar{u} = \overline{\mathcal{U}(u)} = \frac{u_1+u_2+\cdots+u_{s_{max}}}{s_{max}} \quad (8)$$

1D scalogram is used to detect the strength and the position of the edges of the Gaussians.

B. The procedure of CGDwavetlet

In CGDWavelet, the left and the right peak of bimodal histone modification signals are modelled with two separated Gaussians. A nucleosome free region is the region located between the two imaginary Gaussians. Figure 1 summarizes the procedure used for CGDWavelet. First, deep sequencing (ChIP-seq) signals for histone modification are decomposed to multi sub-bands in the wavelet domain. The first Gaussian derivative wavelet is applied to detect the position of the left and the right peak as well as the valley. 1D wavelet scalogram is used to estimate the edges of the imaginary Gaussian. We defined a bimodal peak if the distance between the two Gaussians is larger than 150 base pairs (bps) and smaller than 1500 bps. In the quantification step, we also removed the

bimodal peaks with their heights of the Gaussians were below a certain threshold. After quantification, we received a list of bimodal peaks.

To identify the dynamic nucleosome changes, we investigated the intensity levels of histone modification from two samples. CGDWavelet applies correlation to the obtained Gaussian parameters and classifies epigenetic changes.

Figure 2 demonstrates how CGDWavelet identifies the Gaussian parameters from histone modification signals. The input data (H3K4me1, top panel) is decomposed to the wavelet domain by using Gaussian derivative wavelet (the second panel). Zero-crossing lines detect local maxima that represent the peaks and the valleys of the histone modification signals (the third panel). The length of zero-crossing lines represents the frequency of the corresponding Gaussian peaks. In Figure 2, five blue lines and four green lines were obtained after zero-crossing. Applying 1D wavelet scalogram (the bottom panel), we receive the edges of the Gaussians. After quantification, the lines below a certain threshold are ignored and two bimodal peaks were obtained.

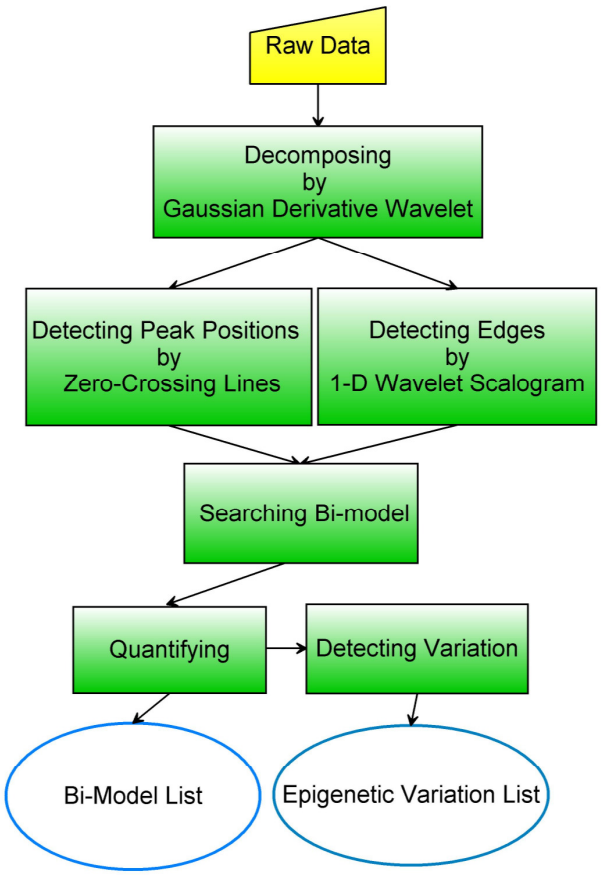

Figure 1. The procedure of CGDWavelet. CGDWavelet is composed a number of steps to identify the parameters of a Gaussian that models a nucleosome. After the entire steps, CGDWavelet produce a list of bimodal peaks and their variations.

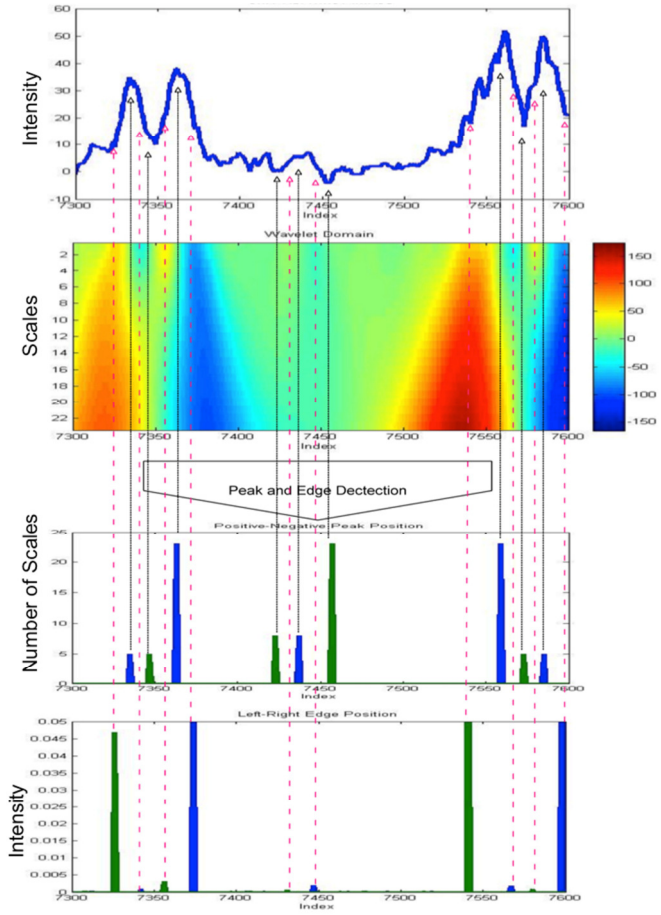

Figure 2. CGDWavelet identifies the Gaussian parameters using zero-crossing and 1D scalogram. Histone modification data (the top panel) is converted into the wavelet domain (the second panel). The vertical axis in the wavelet domain represents the wavelet scale. The zero-crossing lines detect the corresponding peak positions (the third panel). The green lines represent the positive Gaussian peaks (concave) and the blue lines are corresponding to negative Gaussian peaks (convex). The fourth panel shows the detected edges of the bimodal peaks after applying the 1D wavelet scalogram [3].

CGDWavelet calculates a binding score from the obtained Gaussian parameters. A binding score is defined as

a binding score = l(left peak) + l(right peak) -2*l(valley),
where l() is the intensity levels for peaks or valleys.

C. *Clustering*

We applied correlation to the obtained Gaussian parameters to identify epigenetic changes. Based on the correlation coefficient and the signal levels, we defined 4 groups: 1) loss of bimodal peak, 2) only the histone modification levels are changed without any changes in their shape 3) a slight closure of chromatin structure 4) represents a full closure of chromatin structure.

III. RESULT

A. *Identifying bimodal peaks*

Inherited from AWNFR [5], CGDWavelet has the performance as good as AWNFR. Even though the purpose of CGDWavelet is to identify epigenetic variations, we compared the performance of CGDWavelet with the hidden Markov model (HMM) based supervised learning method developed by us previously [9]. For this, we used histone modifications (H3K4me1/2/3) in mESC [24, 25] and evaluated the performance using the known binding sites of 13 TF in mESC [26].

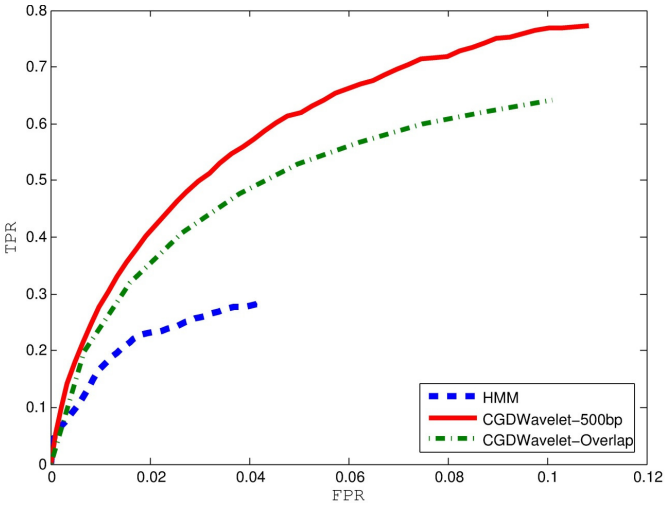

Figure 3. Performance assessment of CGDWavelet against the HMM based method [9]. We defined a prediction as TP when a prediction is located within 500 for the HMM based method and CGDWavelet. We also defined a prediction as TP only when a prediction overlaps with any known TFBS (only for CGDWavelet). CGDWavelet showed superior performance to the HMM-based method.

Using the binding score as a threshold we calculated true predicative rate (TPR) over false predicative rate (FPR) and drew receiver operating characteristic (ROC) curve. A prediction is regarded as a true positive (TP) if it is within 500 bps of any known TFBSs. CGDWavelet showed a better performance compared with the HMM based models [9] (Figure 3). We also used more strict criteria for CGDWavelet by regarding a prediction as true only when the predicted binding site overlaps with any known TFBSs. We found that CGDWavelet outperformed the HMM based approach [9] even when using more strict criteria. This at least suggests that CGDWavelet can be used to identify bimodal peaks.

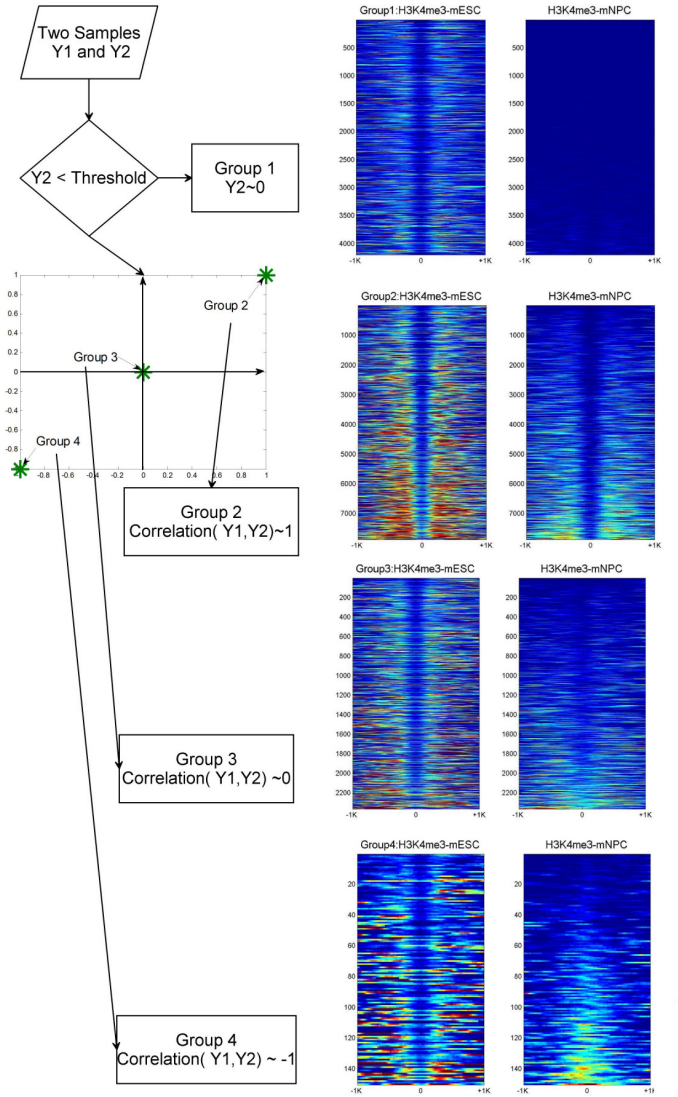

Figure 4. Dynamic epigenetic changes identified using CGDWavelet. We investigated the epigenetic variation using H3K4me3 in mESCs and mNPCs. Group 1: loss of bimodal peak; Group 2: level changes; Group 3: a slight closure of chromatin structure; Group 4: a full closure of chromatin structure.

B. *Classifying epigenetic changes*

Using CGDWavelet we searched for the epigenetic changes. [9]. For this, we used H3K4me1 and H3K4me3 in mESC [24, 25].

We identified total 22,129 and 24,739 bimodal patterns for H3K4me1 and H3K4me3, respectively (FPR 0.025). At these identified bimodal peaks, we calculated correlation coefficient to identify dynamic epigenetic variations. Figure 4 shows the 4 groups identified by CGDWavelet when we used H3K4me3 in mESCs and mNPCs. H3K4me3 signal is lost in Group 1. Group 2 is when the correlation coefficient is close to 1, suggesting that the bimodal pattern does not change in this group. Group 3 and 4 have bimodal peaks to uni-modal peaks.

The chromatin structure is closed slightly in Group 3 (correlation coefficient is close to 0), but significantly in Group 4 (correlation coefficient is close to -1) (Figure 5). This suggests that CGDWavelet effectively identifies the epigenetic changes of the bimodal peaks.

We further investigated the expressions of the associated genes for each group. We found that gene expressions become significantly reduced in all four groups including Group 4 where dynamic nucleosome positions were observed.

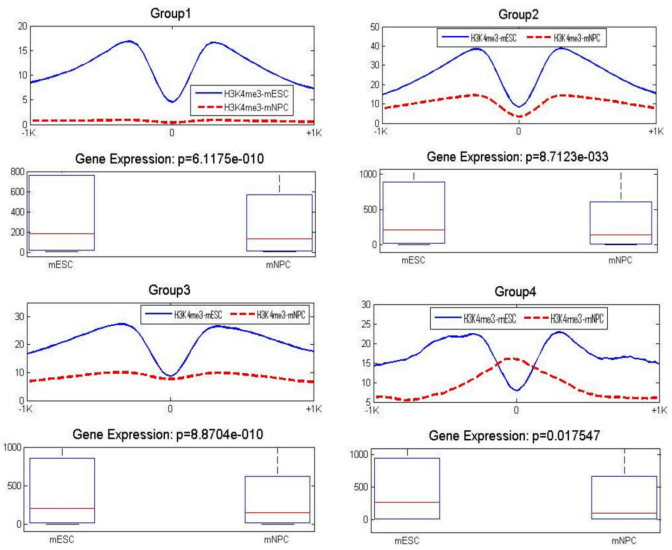

Figure 5. Epigenetic changes and the associated gene expression levels. The average profiles of H3K4me3 for each group identified in Figure 4 are shown. We evaluated the gene expression of the associated genes. The differences of gene expression were significant for all four groups.

C. TFs at the epigenetic changes

We further investigated TFBSs for each group (Table 1). Majority portion of Smad1, Sox2 and Nanog binding sites were observed in Group 1, while other TFs were associated with Group 2. Smad1, Sox2 and Nanog tend to bind at enhancers [10]. This shows that the genomic regions for these enhancer binding factors undergo more epigenetic variation than the other factors in the H3K4me3 enriched regions. Interestingly, majority of Oct4 binding sites were in Group 2, while Sox2, which has been known to form a complex [26], has more binding sites for Group 1.

For the bimodal peaks identified using H3K4me1, majorities were belonged to Group 1 for all factors (Table 2), confirming again that H3K4me1, a histone modification marker for enhancer, is more dynamic than H3K4me3 [27]. Also, the percentages of binding sites to each group were similar in H3K4me1. More than 70% of TFBSs belonged to Group 1 for all TFs.

Table 1. The number of TFBSs belong to each group. H3K4me3, a marker for active promoter, is used.

TFs	Number of TFBSs	Group 1	Group 2	Group 3	Group 4
Smad1	86	70%	17%	13%	0
Sox2	335	50%	37%	12%	1%
Nanog	498	45%	44%	10%	0%
STAT3	399	40%	48%	11%	1%
Esrrb	1804	36%	51%	13%	0%
CTCF	937	35%	50%	15%	1%
Tcfcp2l1	2392	29%	56%	14%	0%
Oct4	592	27%	60%	13%	0%
Klf4	2610	22%	67%	11%	1%
E2f1	5839	19%	68%	12%	1%
Zfx	3088	13%	76%	10%	1%
n-Myc	2908	13%	75%	12%	1%
c-Myc	1508	11%	75%	13%	0%

Table 2. The number of TFBSs belong to each group. H3K4me1, a marker for enhancer, is used.

TFs	Number of TFBSs	Group 1	Group 2	Group 3	Group 4
Smad1	160	83%	10%	7%	1%
Klf4	472	80%	10%	10%	1%
Nanog	577	79%	10%	11%	1%
E2f1	725	78%	12%	9%	1%
Esrrb	799	78%	10%	11%	1%
Sox2	379	77%	11%	11%	1%
Tcfcp2l1	751	76%	13%	10%	1%
STAT3	191	75%	12%	13%	1%
c-Myc	53	74%	15%	11%	0%
Oct4	261	74%	11%	12%	2%
Zfx	262	74%	15%	11%	0%
n-Myc	111	73%	14%	12%	1%
CTCF	178	71%	16%	12%	1%

IV. CONCLUSION

As we accumulate more histone modification data than ever, computational approaches to exploit their landscapes will be of great value. We present an approach to detect epigenomic changes at gene regulatory regions. For this we used active histone modification marks. Because our interest lies in gene

regulatory regions, we focused on the change of the epigenetic shape, which is clearly different from other approaches that only consider the changes in the level of epigenetic signals [19-21]. In contrast, CGDWavelet cannot be used to other marks enriched at gene body (such as H3K36me3) or repressed regions (such as H3K27me3 or H3K20me3).

Changes in active histone marks (H3K4me1 and H3K4me3) were studied genome-widely as well as in association with TF binding. More specifically, we searched for the bimodal peaks with epigenetic changes. We observed that TFBSs associated with the H3K4me1 bimodal peaks were more variable than the H3K4me3 bimodal peaks, confirming previous observation that enhancers are more cell-type specific [27]. We also observed that the changes of chromatin structure in the promoter region affect gene expression. The expression levels were higher when the chromatin structure was open. Previous study shows that both epigenetic pattern as well as its level is important to predict gene expression [28]. Our results also suggest that the position of nucleosome is important for gene regulation.

Our strategy identified cell-type specific enrichment of epigenetic signals (Group 1) and dynamic nucleosome positions (Group 4). As we applied CGDWavelet to the histone modification data during neural development, it may be natural that most of the TFBSs were belonged to Group 1 or 2 because there are cell-type specific as well as common enrichment of histone modification. As a result, we did not find many TFs for Group 4. However, Group 4 will be useful in investigating the dynamic nucleosome positions affected by external signals, where epigenetic landscapes do not change drastically to commit to a certain cell-type.

ACKNOWLEDGMENT

This work is supported by R21-DK098769 and P30-DK19525 from the National Institutes of Diabetes, and Digestive and Kidney Diseases and the Diabetes Research Center at the University of Pennsylvania.

References

[1] T. Kouzarides, "Chromatin modifications and their function," *Cell,* vol. 128, pp. 693-705, Feb 23 2007.

[2] B. E. Bernstein, A. Meissner, and E. S. Lander, "The mammalian epigenome," *Cell,* vol. 128, pp. 669-81, Feb 23 2007.

[3] N. Nguyen, H. Huang, S. Oraintara, and A. Vo, "Mass spectrometry data processing using zero-crossing lines in multi-scale of Gaussian derivative wavelet," *Bioinformatics,* vol. 26, pp. i659-65, Sep 15 2010.

[4] A. K. Maunakea, I. Chepelev, and K. Zhao, "Epigenome mapping in normal and disease States," *Circ Res,* vol. 107, pp. 327-39, Aug 6 2010.

[5] N. Nguyen, A. Vo, and K. J. Won, "A wavelet-based method to exploit epigenomic language in the regulatory region," *Bioinformatics,* Feb 27 2014.

[6] H. H. He, C. A. Meyer, H. Shin, S. T. Bailey, G. Wei, Q. Wang, Y. Zhang, K. Xu, M. Ni, M. Lupien, P. Mieczkowski, J. D. Lieb, K. Zhao, M. Brown, and X. S. Liu, "Nucleosome dynamics define transcriptional enhancers," *Nat Genet,* vol. 42, pp. 343-7, Apr 2010.

[7] Z. Wang, C. Zang, J. A. Rosenfeld, D. E. Schones, A. Barski, S. Cuddapah, K. Cui, T. Y. Roh, W. Peng, M. Q. Zhang, and K. Zhao, "Combinatorial patterns of histone acetylations and methylations in the human genome," *Nat Genet,* vol. 40, pp. 897-903, Jul 2008.

[8] R. D. Hawkins, G. C. Hon, L. K. Lee, Q. Ngo, R. Lister, M. Pelizzola, L. E. Edsall, S. Kuan, Y. Luu, S. Klugman, J. Antosiewicz-Bourget, Z. Ye, C. Espinoza, S. Agarwahl, L. Shen, V. Ruotti, W. Wang, R. Stewart, J. A. Thomson, J. R. Ecker, and B. Ren, "Distinct epigenomic landscapes of pluripotent and lineage-committed human cells," *Cell Stem Cell,* vol. 6, pp. 479-91, May 7 2010.

[9] K. J. Won, I. Chepelev, B. Ren, and W. Wang, "Prediction of regulatory elements in mammalian genomes using chromatin signatures," *BMC Bioinformatics,* vol. 9, p. 547, 2008.

[10] K. J. Won, B. Ren, and W. Wang, "Genome-wide prediction of transcription factor binding sites using an integrated model," *Genome Biol,* vol. 11, p. R7, Jan 22 2010.

[11] J. Ernst, H. L. Plasterer, I. Simon, and Z. Bar-Joseph, "Integrating multiple evidence sources to predict transcription factor binding in the human genome," *Genome Res,* vol. 20, pp. 526-36, Apr 2010.

[12] H. A. Firpi, D. Ucar, and K. Tan, "Discover regulatory DNA elements using chromatin signatures and artificial neural network," *Bioinformatics,* vol. 26, pp. 1579-86, Jul 1 2010.

[13] P. C. FitzGerald, D. Sturgill, A. Shyakhtenko, B. Oliver, and C. Vinson, "Comparative genomics of Drosophila and human core promoters," *Genome biology,* vol. 7, p. R53, 2006.

[14] K. J. Polach and J. Widom, "Mechanism of protein access to specific DNA sequences in chromatin: a dynamic equilibrium model for gene regulation," *Journal of molecular biology,* vol. 254, pp. 130-49, Nov 24 1995.

[15] L. A. Mirny, "Nucleosome-mediated cooperativity between transcription factors," *Proceedings of the National Academy of Sciences of the United States of America,* vol. 107, pp. 22534-9, Dec 28 2010.

[16] Y. Zhang, H. Shin, J. S. Song, Y. Lei, and X. S. Liu, "Identifying positioned nucleosomes with epigenetic marks in human from ChIP-Seq," *BMC Genomics,* vol. 9, p. 537, 2008.

[17] S. Heinz, C. Benner, N. Spann, E. Bertolino, Y. C. Lin, P. Laslo, J. X. Cheng, C. Murre, H. Singh, and C. K. Glass, "Simple combinations of lineage-

determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities," *Mol Cell,* vol. 38, pp. 576-89, May 28 2010.

[18] A. Polishko, N. Ponts, K. G. Le Roch, and S. Lonardi, "NORMAL: accurate nucleosome positioning using a modified Gaussian mixture model," *Bioinformatics,* vol. 28, pp. i242-9, Jun 15 2012.

[19] H. Xu, C. L. Wei, F. Lin, and W. K. Sung, "An HMM approach to genome-wide identification of differential histone modification sites from ChIP-seq data," *Bioinformatics,* vol. 24, pp. 2344-9, Oct 15 2008.

[20] L. Shen, N. Y. Shao, X. Liu, I. Maze, J. Feng, and E. J. Nestler, "diffReps: detecting differential chromatin modification sites from ChIP-seq data with biological replicates," *PloS one,* vol. 8, p. e65598, 2013.

[21] H. Xu and W. K. Sung, "Identifying differential histone modification sites from ChIP-seq data," *Methods in molecular biology,* vol. 802, pp. 293-303, 2012.

[22] P. Zhang, Y. Liu, Z. Y. Liu, Y. Shen, and J. H. Yu, "[Analysis of wavelet scalogram of blood flow ultrasonic Doppler signal]," *Hang tian yi xue yu yi xue gong cheng = Space medicine & medical engineering,* vol. 13, pp. 119-23, Apr 2000.

[23] S. Mallat, *Wavelet Tour of Signal Processing - The Sparse Way*: Academic Press, 2009.

[24] T. S. Mikkelsen, M. Ku, D. B. Jaffe, B. Issac, E. Lieberman, G. Giannoukos, P. Alvarez, W. Brockman, T. K. Kim, R. P. Koche, W. Lee, E. Mendenhall, A. O'Donovan, A. Presser, C. Russ, X. Xie, A. Meissner, M. Wernig, R. Jaenisch, C. Nusbaum, E. S. Lander, and B. E. Bernstein, "Genome-wide maps of chromatin state in pluripotent and lineage-committed cells," *Nature,* vol. 448, pp. 553-60, Aug 2 2007.

[25] A. Meissner, T. S. Mikkelsen, H. Gu, M. Wernig, J. Hanna, A. Sivachenko, X. Zhang, B. E. Bernstein, C. Nusbaum, D. B. Jaffe, A. Gnirke, R. Jaenisch, and E. S. Lander, "Genome-scale DNA methylation maps of pluripotent and differentiated cells," *Nature,* vol. 454, pp. 766-70, Aug 7 2008.

[26] X. Chen, H. Xu, P. Yuan, F. Fang, M. Huss, V. B. Vega, E. Wong, Y. L. Orlov, W. Zhang, J. Jiang, Y. H. Loh, H. C. Yeo, Z. X. Yeo, V. Narang, K. R. Govindarajan, B. Leong, A. Shahab, Y. Ruan, G. Bourque, W. K. Sung, N. D. Clarke, C. L. Wei, and H. H. Ng, "Integration of external signaling pathways with the core transcriptional network in embryonic stem cells," *Cell,* vol. 133, pp. 1106-17, Jun 13 2008.

[27] N. D. Heintzman, G. C. Hon, R. D. Hawkins, P. Kheradpour, A. Stark, L. F. Harp, Z. Ye, L. K. Lee, R. K. Stuart, C. W. Ching, K. A. Ching, J. E. Antosiewicz-Bourget, H. Liu, X. Zhang, R. D. Green, V. V. Lobanenkov, R. Stewart, J. A. Thomson, G. E. Crawford, M. Kellis, and B. Ren, "Histone modifications at human enhancers reflect global cell-type-specific gene expression," *Nature,* Mar 18 2009.

[28] Y. Chen, M. Jorgensen, R. Kolde, X. B. Zhao, B. Parker, E. Valen, J. Y. Wen, and A. Sandelin, "Prediction of RNA Polymerase II recruitment, elongation and stalling from histone modification data," *Bmc Genomics,* vol. 12, Nov 30 2011.