

METHODOLOGY ARTICLE

Open Access

# Analysis method of epigenetic DNA methylation to dynamically investigate the functional activity of transcription factors in gene expression

Weixing Feng<sup>\*</sup>, Zengchao Dong, Bo He and Kejun Wang

## Abstract

**Background:** DNA methylation is a fundamental component of epigenetic modification, which is intimately involved in the regulation of gene expression. One important DNA methylation pathway reduces the abilities of transcription factors to bind to gene promoter regions. Although many experiments have been designed to measure genome-wide DNA methylation levels at high resolution, the meaning of these different DNA methylation levels on transcription factor binding abilities remains poorly understood. We have, therefore, developed a method to quantitatively explore the extent to which DNA methylation levels can significantly reduce or even abolish the binding of certain transcription factors, resulting in reduced or non-expression of flanking genes. This method allows transcription factors that are functionally active in gene expression to be investigated.

**Results:** The method is based on a general model that depicts the relationship between DNA methylation and transcription factor binding ability based on intrinsic component properties, and the model parameters can be optimized through relative analysis of recognized transcription factor binding status and gene expression profiling. With fixed models, transcription factors functionally active in the regulation of gene expression and affected by epigenetic DNA methylation can be identified and subsequently confirmed. The method identified eleven apparently functionally active transcriptional factors in SH-SY5Y neuroblastoma cells.

**Conclusions:** Compared with gene regulatory elements, epigenetic modifications are able to change to dynamically respond to signals from physical, biological and social environments. Our proposed method is therefore designed to provide a dynamic assessment to investigate functionally active transcription factors. With the information deduced from our method, we can predict transcription factor binding status in promoter regions to further investigate how a particular gene is regulated by a specific group of transcription factors organized in a particular pattern. This will be helpful in the diagnosis and development of treatment for numerous diseases, including cancer. Although the method only investigates DNA methylation, it has the potential to be applied to more epigenetic factors, such as histone modification.

**Keywords:** Method, Epigenetics, DNA methylation, Activity, Transcription factor, Gene expression

## Background

Epigenetics is defined as the study of heritable modifications to gene function that occur without alterations in DNA sequences. Epigenetic modifications consist mainly of DNA methylation, histone modifications, chromatin reconstruction, and expression of non-coding RNA. Epigenetic modifications are widely recognized to regulate

tissue-specific gene expression, genomic imprinting and X-chromosome inactivation. In addition, the key role of epigenetic modifications during cellular differentiation, development and organogenesis has been highlighted by the identification of many epigenetic biomarkers in human diseases [1,2], such as neuroblastic tumors [3].

The occurrence of many human cancers results from the accumulation of both genetic and epigenetic alterations. While genetic alterations are nearly impossible to reverse, epigenetic alterations can dynamically respond to signals from physical, biological and social environments

\* Correspondence: fengweixing@hrbeu.edu.cn  
Pattern Recognition and Intelligent System Institute, Automation College,  
Harbin Engineering University, Harbin, Heilongjiang, China

[4-6]. This characteristic confers the importance of epigenetic research in various cellular processes, especially in gene expression regulation. Although epidemiological data provide evidence that there are direct interactions between epigenetic modifications and the environment to influence gene expression, the mechanism of epigenetic induced modulations of gene expression is still poorly understood.

Regulation of gene expression by transcription factors is a fundamental mechanism. Through the interplay with transcription factors, epigenetic modification such as DNA methylation is able to regulate gene expression [7-11]. For example, high methylation levels in promoter regions always weaken the binding ability of associated transcription factors and cause reduced expression of adjacent genes [12,13]. Although there are many qualitative observations about the effect of DNA methylation on gene regulation, few methods have been developed to assess the effect in a measurable way. Here, we propose a method to evaluate how each transcription factor affects gene expression under a specific pattern of epigenetic DNA methylation levels, which is then used to determine the functional activity of the transcription factor. We describe a general model of how epigenetic DNA methylation affects transcription factor binding ability where several model parameters provide sufficient freedom for different circumstances. Through the relative analysis of recognized transcription factor binding status and gene expression profiling, a model for each transcription factor can be fixed with concrete parameter values. Then, with the deduced models, transcription factors affected by DNA methylation and functionally active in gene expression can be investigated. The proposed method has the capacity to dynamically reflect functions of transcription factors in a temporal and spatial manner.

## Methods

In addition to gene sequence-driven gene regulatory mechanisms, epigenetic modifications, such as DNA methylation, also participate in the regulation of gene expression induced by signals from the environment. Here, based on genome-wide DNA methylation profiling in gene promoter regions, we present a method to investigate transcription factors that are affected by DNA methylation and that are functionally active in gene expression.

### Transcription factor match score

As a functional protein, a transcription factor has the intrinsic tendency to combine with specific DNA sequences, and we define a value termed 'transcription factor match score' to evaluate such binding ability for

each transcription factor in the promoter region of each gene. In the TRANSFAC database produced by BIOBASE, position weight matrices (PWMs) for every transcription factor are provided. In these matrices, each row consists of four weights representing different capabilities to combine with nucleotides A, C, G and T, respectively. Using these PWMs, each gene promoter region can be scanned nucleotide by nucleotide with a smooth window to compute transcription factor match scores.

For the  $i$ th transcription factor with a motif length of  $L$ , a match value at the  $k$ th putative binding site in promoter region of the  $j$ th gene can be calculated as  $A_{ijk}$ .

$$A_{ijk} = \sum_{l=1}^L a_{jkl} w_{il}^T \quad (1)$$

where  $a_{jkl}$  is the nucleotide (A, C, G, T) at the  $l$ th position in the  $k$ th possible binding site in promoter region of the  $j$ th gene, and  $w_{il}$  is the  $l$ th nucleotide in the match weights row vector in the PWM of the  $i$ th transcription factor. So, suppose the length of the promoter region of the  $j$ th gene is  $N$ ,  $N-L+1$  match values can be calculated and the maximal value is adopted as the match score,  $S_{ij}$ , to reflect the binding ability of the  $i$ th transcription factor in the  $j$ th gene promoter region.

$$S_{ij} = \max A_{ijk} \quad (2)$$

Hence, for one transcription factor, a collection of match scores can be calculated with respect to every gene promoter region.

Although match scores can approximate the opportunity for a transcription factor to bind to a gene promoter region, it is also meaningful to determine a threshold for match scores to evaluate whether the transcription factor binds and regulates the transcription of specific genes.

### Transcription factor match score threshold

As described in the method proposed by Hertzberg [14], for a given transcription factor, a  $Z$ -score, which considers the relationship between transcription factor match scores and gene expression levels, was calculated to infer the match score threshold.

Suppose there are  $n$  genes in a cell. Let  $g_1, \dots, g_n$  be their log expression values, which would follow normal distribution with an average of  $\mu$  and a standard deviation of  $\sigma$ . If a threshold,  $t$ , is set for match scores of a transcription factor, there will be a subgroup of  $k$  genes,  $G_i$ , whose match scores are greater than the threshold  $t$ , and  $G_i$  are assumed to be targets of the transcription factor. The log expression values of these selected genes,  $G_i$ , also approximately follow normal distribution for a

large number of elements in  $G_i$ . The  $Z$ -score can then be calculated.

$$Z(TF, G^i) = \left( \frac{1}{k} \sum_{j=1}^k g_{ij} - \mu \right) / (\sigma / \sqrt{k}) \quad (3)$$

The  $Z$ -score reflects the extent to which average expression of the selected target genes differ from average expression of all genes. In other words, a larger absolute  $Z$ -score value means a higher relationship between transcription factor match scores and expression of selected genes, and that these genes are more likely to be regulated by the same transcription factor. Therefore, with different thresholds for transcription factor match scores, we can obtain different groups of transcription factor target genes and subsequently different  $Z$ -score values. Finally, the best threshold can be determined when the maximal  $Z$ -score value (if positive) or the minimal  $Z$ -score value (if negative) is found, where the corresponding  $Z$ -score for the particular transcription factor is called  $Z_m$ .

However, without considering the effects of epigenetic modifications, the match score defined above only considers DNA sequences to decide whether a transcription factor binds and regulates the expression of certain genes. This undoubtedly makes subsequent  $Z$ -score values inaccurate in the evaluation of transcription factor binding status in gene promoter regions. Hence, we have improved the calculation of the transcription factor match score by adding the effect of epigenetic modifications. However, among the many epigenetic modifications, only DNA methylation was considered because of the requirement for high precision and high resolution data.

### General model of DNA methylation effect

DNA methylation is known to repress transcription factor binding ability [15-23]; therefore, we designed a general model to describe such an effect, where a nonlinear  $S$ -function is adopted to normalize the effect between 0 and 1. The model consists of two parts. The first sense part uses an inverse  $S$ -function (Equation 4) to accurately depict the DNA methylation effect. In the equation,  $M_{jk}$  is the methylation level at the  $k$ th putative transcription factor binding site in the promoter region of the  $j$ th gene, and  $C_i$  and  $S_i$  are parameters of the model for the  $i$ th transcription factor.

$$E_{ijk1} = \frac{e^{-((M_{jk}-C_i)/S_i)}}{1 + e^{-((M_{jk}-C_i)/S_i)}} \quad (4)$$

Two biological observations are considered here. A larger methylation level results in reduced transcription factor binding ability in a monotonic way and *vice versa*.

Next, the sensitivity of the methylation effect on transcription factor binding ability is not the same at different methylation levels. When the methylation level is quite large or small, the effect tends to be saturating to 0 or 1 and insensitive to a change in methylation level. In contrast, the effect will be sensitive to change when the methylation level is around a median value. Here, an inverse  $S$ -function is capable of fitting such a relationship, which is shown in Figure 1 (solid line, the parameters are assumed as  $C=0$  and  $S=0.1$ ). In Figure 1, the methylation level is on the  $X$  axis and the suppression of transcription factor binding ability is on the  $Y$  axis. In the general model, two parameters,  $C$  and  $S$ , of the inverse  $S$ -function are adjustable and can be tuned for different transcription factors in a specific cell.

To increase sensitivity of the method, we also propose a second part to the general model to depict the effect of DNA methylation in an antisense way (Figure 1, dashed line), where a normal  $S$ -function was used (Equation 5). In the model, a large methylation level was assumed to impact less on transcription factor binding ability and *vice versa*.

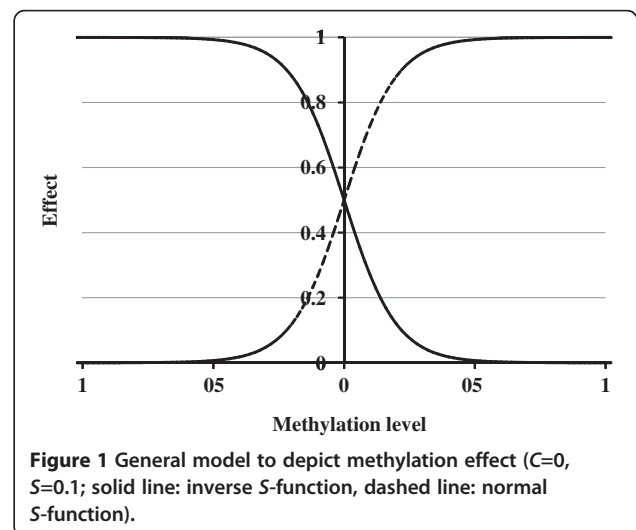
$$E_{ijk2} = \frac{1}{1 + e^{-((M_{jk}+C_i)/S_i)}} \quad (5)$$

### Transcription factor binding score

With consideration of the DNA methylation effect, the binding ability of the  $i$ th transcription factor in the  $j$ th gene promoter region can be modified as binding score  $B_{ij}$  from match score  $M_{ij}$ .

$$B_{ij} = \max_k (A_{ijk} \times E_{ijk}) \quad (6)$$

where  $A_{ijk}$  is the sequences match value of the  $i$ th transcription factor and  $E_{ijk}$  is the effect of DNA methylation



on the binding ability of the  $i$ th transcription factor at the  $k$ th putative binding site in the promoter region of the  $j$ th gene.

Similar to the transcription factor match score, by threshold analysis of the transcription factor binding score, a maximal  $Z$ -score value (if positive) or a minimal  $Z$ -score value (if negative), known as  $Z_m$ , can also be calculated based on the relative analysis of transcription factor binding scores and gene expression profiles. However, in contrast to only one  $Z_m$  value based on the match score, there are many  $Z_m$  values for a transcription factor when different compositions of parameters  $C$  and  $S$  are selected in the model to calculate different binding scores. Then, when parameters  $C$  and  $S$  of the model are fixed to obtain an optimized  $Z_m$  value, the effect of methylation on transcription factor binding ability can be quantitatively determined.

### Functionally active transcription factors

According to different ways of describing DNA methylation effects on transcription factor binding ability, three  $Z_m$  values can be calculated to investigate functionally active transcription factors. Without considering a DNA methylation effect,  $Z_{m-o}$  is computed when transcription factor match scores are adopted. In contrast, with the consideration of a DNA methylation effect using our proposed model,  $Z_{m-p}$  is analyzed from transcription factor binding scores from the sense orientation and  $Z_{m-q}$  is calculated from transcription factor binding scores from the antisense orientation. Furthermore, with different compositions of model parameters, a group of  $Z_{m-p}$  and  $Z_{m-q}$  values can be calculated for each transcription factor. Then, if absolute  $Z_{m-p}$  values are found to be obviously larger than the absolute  $Z_{m-o}$  value and absolute  $Z_{m-q}$  values are always less than the absolute  $Z_{m-o}$  value, the transcription factor is considered to be affected by DNA methylation and functionally active.

## Results and discussion

### Materials

Rett syndrome, a condition frequently seen in cases of developed neuroblastoma, is caused by abnormal interactions between binding proteins and methylated DNA in promoter regions. To evaluate the utility of our proposed method for the investigation of active transcription factors with respect to DNA methylation, a dataset from the SH-SY5Y thrice-cloned neuroblastoma cell line (made by ATCC) was used. As indicated, the dataset includes two parts:

Part 1: DNA methylation levels in promoter regions of SH-SY5Y neuroblastoma cells, assayed with the NimbleGen-1500b-Promoter-Array in the MeDIP

experiment, were retrieved (GSE9568). The promoter regions covered by the array are 1200 bps upstream and 300 bps downstream of gene transcriptional start sites. Log<sub>2</sub>-ratios of the Cy5-labeled test sample versus the Cy3-labeled reference sample were calculated to represent DNA methylation levels. Then, methylation levels of every specific transcription factor binding site (about 10 bps) in all promoter regions were calculated using the Batman algorithm [24].

Part 2: Gene expression levels in the same SH-SY5Y cell line under the same conditions, measured using Affymetrix-HG-U133-plus2.0 GeneChips, were retrieved (GSE4600). Using human RefSeq gene annotations downloaded from the server at UCSC, 10065 available gene expression levels were identified. To enhance observation of the interaction between DNA methylation and transcription factors in the regulation of gene expression, we filtered out genes with very low expression levels.

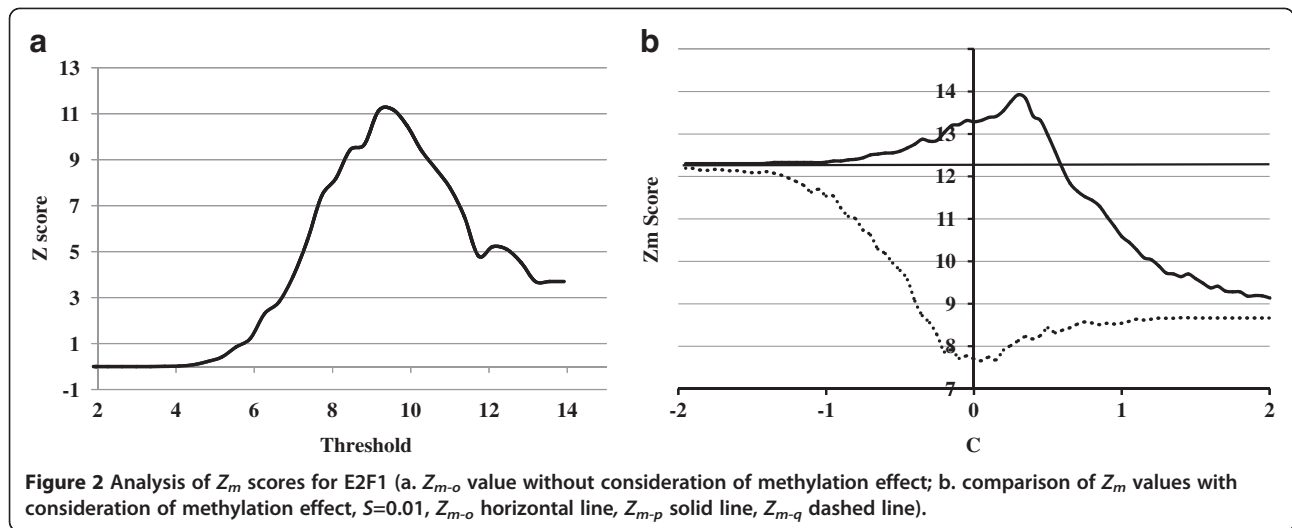
### Differences in transcription factor binding abilities with and without consideration of a methylation effect

PWMs of 459 human transcription factors were extracted from the TRANSFAC database. With these transcription factor PWMs, match scores for each transcription factor in all gene promoter regions were calculated on human DNA sequences from UCSC. Then, based on DNA methylation levels and gene expression data in SH-SY5Y cells, for each transcription factor we calculated the  $Z_{m-o}$  value, without consideration of the DNA methylation effect, and  $Z_{m-p}$  and  $Z_{m-q}$  values, with consideration of DNA methylation effect, using our proposed sense and antisense models, respectively. Differences among  $Z_{m-o}$  values and extreme  $Z_{m-p}$  values of all transcription factors, with and without consideration of a methylation effect, were found to be significant (Wilcoxon, P-Value < 2.2e-16), and showed a different distribution of extreme  $Z_{m-p}$  values compared with that of  $Z_{m-o}$  values when considering a DNA methylation effect on transcription factor binding ability.

### Investigation of functionally active transcription factors

Among 459 human transcription factors, E2F1 was reported to be rich in SH-SY5Y cells and to be affected by DNA methylation [25]; therefore, first we show the analysis process of E2F1 in detail.

Neglecting the effect of methylation, we computed and plotted different  $Z$ -scores while adjusting the E2F1 match score threshold (Figure 2a). The positive  $Z_m$  values in Figure 2a mean that E2F1 is an active factor in SH-SY5Y cells, which is in accordance with previous studies [25]. The maximal  $Z$ -score value ( $Z_{m-o}$ ) was 12.30 at a match score threshold 9.57.



The effect of methylation with the sense part of model was then considered and  $Z_{m-p}$  values were calculated along with adjustment of model parameters, and the extreme  $Z_{m-p}$  value was found as 13.923 when parameters of the model were selected as  $C=-0.25$  and  $S=0.01$ . The searching process of  $Z_{m-p}$  values is shown in Figure 2b (solid line).

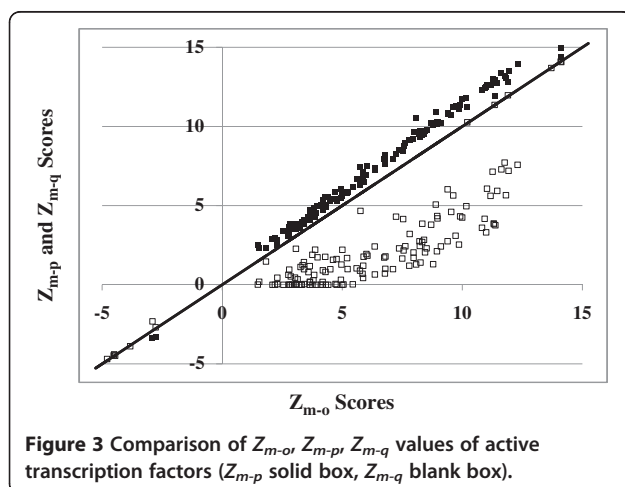
In Figure 2b, the X axis is the center  $C$  of the general model, from -2 to 2 and stepped by 0.05, and the Y axis is the corresponding  $Z_m$  values (as steepness  $S$  of the general model contributes little to the effect of the model compared with  $C$ , to clearly exhibit the searching process of  $Z_m$  values,  $S$  is set at a fixed value of 0.01). The horizontal solid line at 12.30 indicates the  $Z_{m-o}$  value without consideration of a methylation effect.

In Figure 2b, while increasing the value of  $C$  from -2 to gradually strengthen the methylation effect, the  $Z_{m-p}$  value begins to rise and soon becomes greater than the  $Z_{m-o}$  value. This means a more reasonable result is

obtained when considering a methylation effect on transcription factor binding ability with the sense part of model. The  $Z_{m-p}$  value reaches its highest value at 13.92 (13% higher than the  $Z_{m-o}$  value) when  $C$  is 0.25. After that, the  $Z_{m-p}$  value drops rapidly when the effect of DNA methylation is further increased.

We also used the antisense part of the model to analyze  $Z_{m-q}$  values along with adjustment of model parameters. The result is shown in Figure 2b (dashed line). While increasing the value of  $C$  from -2 to gradually weaken the methylation effect, the  $Z_{m-q}$  value reduces and remains lower than the  $Z_{m-o}$  value. This means the antisense part depicts the effect of methylation on E2F1 in an incorrect way. According to our proposed method, we determined that E2F1 was affected by DNA methylation and was functionally active in gene expression in SH-SY5Y cells.

Active transcription factors in SH-SY5Y cells were then investigated. After analysis of distributions of  $Z_{m-o}$ ,



**Table 1** Top 10 positive transcription factors in SH-SY5Y cells

No	AC	Name	$Z_{m-o}$	Extreme $Z_{m-p}$	C	S
1	M00938	E2F1	12.300	13.923	0.25	0.01
2	M00716	ZF5	11.760	13.189	0	0.01
3	M00189	AP2	9.991	11.714	-0.2	0.1
4	M00196	Sp1	9.908	11.351	-0.05	0.01
5	M00466	HIF1	9.620	11.213	0.15	0.01
6	M00332	Whn	8.892	10.885	0.1	0.01
7	M00778	AhR	9.385	10.768	-0.05	0.01
8	M00801	CREB	8.075	10.562	-0.1	0.05
9	M00245	Egr3	8.965	10.307	-0.1	0.1
10	M00982	KROX	8.701	10.194	0	0.01



**Table 2 Negative transcription factor in SH-SY5Y cells**

No	AC	Name	$Z_{m-o}$	Extreme $Z_{m-p}$	C	S
1	M00526	GCNF	-2.906	-3.390	0.35	0.01

$Z_{m-p}$  and  $Z_{m-q}$  values of all transcription factors, we adopted two standards to identify active transcription factors. First, an absolute extreme  $Z_{m-p}$  value was at least 10% (90<sup>th</sup> percentile of ratio values) greater than the absolute  $Z_{m-o}$  value. Second, absolute  $Z_{m-q}$  values were always less than the absolute  $Z_{m-o}$  value. The  $Z_{m-o}$ ,  $Z_{m-p}$  and  $Z_{m-q}$  values of identified active transcription factors are shown in Figure 3. In Figure 3, the X axis is  $Z_{m-o}$  values without consideration of a methylation effect and the Y axis is extreme  $Z_m$  values with consideration of a methylation effect (extreme  $Z_{m-p}$  values are filled boxes and extreme  $Z_{m-q}$  values are unfilled boxes).

Among the identified active transcription factors, the top 10 functionally positive transcription factors (removal of redundant transcription factors) ranked by extreme  $Z_{m-p}$  values in descending order are listed in Table 1. Besides these positive transcription factors, one functionally negative transcription factor, germ cell nuclear factor (GCNF), is listed in Table 2. Comparisons of  $Z_{m-o}$ ,  $Z_{m-p}$  and  $Z_{m-q}$  values of first two transcription factors (ZF5 and AP2) in Table 1 are shown in Figure 4, and those comparisons of other factors in Tables 1 and 2 are shown in Additional file 1. Similar to E2F1, by calculating and analyzing  $Z_m$  values with and without consideration of a methylation effect, the listed transcription factors were found to be in the same pattern as E2F1. Among these active transcription factors, although E2F1, AP2, Sp1 and CREB exist ubiquitously in many tissues, other transcription factors, ZF5, HIF1, AhR and Egr3 were found to specifically exist in SH-SY5Y neuroblastoma

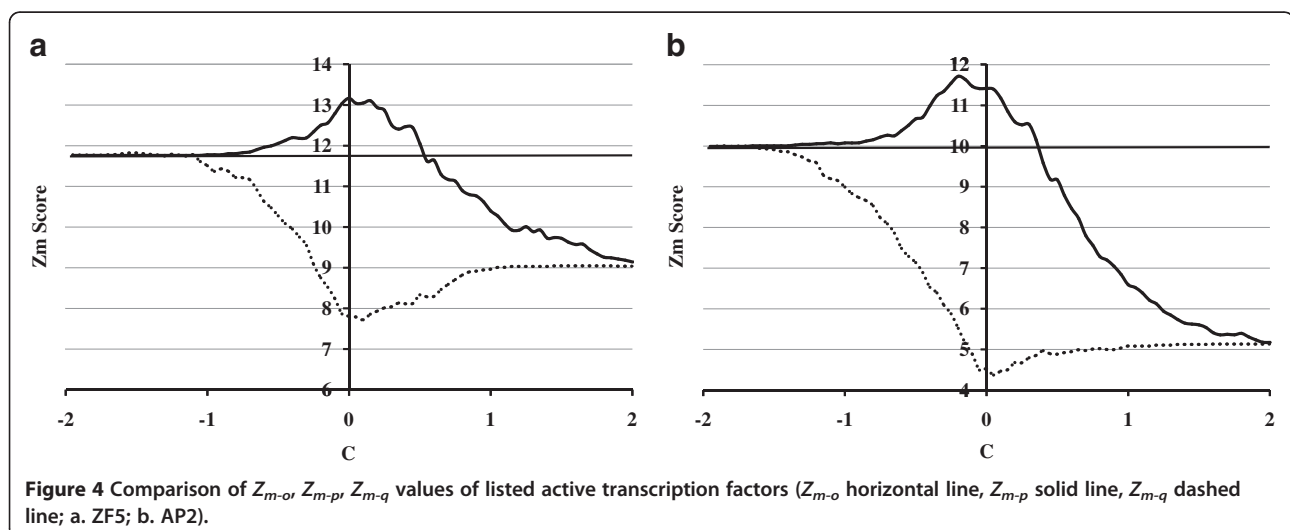
cells [26]. Binding of all these transcription factors are found to be affected by DNA methylation [16-21]. Meanwhile, the negative GCNF is also a sequence-specific repressor of transcription through interaction with methylated DNA [22] and functions in neural differentiation [23].

In the method described here, methylation effects on binding abilities of different transcription factors need to be described for each transcription factor; a model is designed with particular independent parameters for each transcription factor. In future research, we will improve the performance of the method by considering transcription factor clustering and multiple transcription factors acting at their binding sites in modules.

### Conclusions

In this study, we have proposed a method to detect active transcription factors in specific cell types by analyzing the interactions between epigenetic methylation patterns in gene promoter regions and the expected binding of transcription factors. In the method, we designed a general model to quantitatively analyze the effect of methylation to suppress transcription factor binding ability in promoter regions, where an inverse S-function was adopted to depict the effect of methylation and the model parameters were fixed through calculation of the relationship between transcription factor binding scores in promoter regions and gene expression levels. Based on the model, the case analysis of data from a neuroblastoma cell line successfully showed that 11 transcription factors were obviously affected by methylation of promoter regions and were functionally active in gene expression.

Besides detection of active transcription factors, information deduced from the model can indicate transcription factor binding status in promoter regions to further



investigate how a particular gene is regulated by a specific group of transcription factors organized in a particular pattern. This should be helpful for diagnosis and for the development of treatments for numerous diseases, including various cancers. The prediction of transcription factor binding sites produces many false positives; however, by combining static genetic and dynamic epigenetic information together, our proposed approach is capable of effectively decreasing the false positive rate.

So far, we have only considered DNA methylation in the proposed method because of the requirement for high precision and high resolution data. But, the method has the potential to consider more epigenetic factors, such as histone modifications, when the quality of data improves with the development of experimental technology.

## Additional file

**Additional file 1: FigureS1.** Comparison of Zm-o, Zm-p, Zm-q values of Sp1 (Zm-o horizontal line, Zm-p solid line, Zm-q dashed line). FigureS2. Comparison of Zm-o, Zm-p, Zm-q values of HIF1 (Zm-o horizontal line, Zm-p solid line, Zm-q dashed line). FigureS3. Comparison of Zm-o, Zm-p, Zm-q values of Whn (Zm-o horizontal line, Zm-p solid line, Zm-q dashed line). FigureS4. Comparison of Zm-o, Zm-p, Zm-q values of AhR (Zm-o horizontal line, Zm-p solid line, Zm-q dashed line). FigureS5. Comparison of Zm-o, Zm-p, Zm-q values of CREB (Zm-o horizontal line, Zm-p solid line, Zm-q dashed line). FigureS6. Comparison of Zm-o, Zm-p, Zm-q values of Egr3 (Zm-o horizontal line, Zm-p solid line, Zm-q dashed line). FigureS7. Comparison of Zm-o, Zm-p, Zm-q values of KROX (Zm-o horizontal line, Zm-p solid line, Zm-q dashed line). FigureS8. Comparison of Zm-o, Zm-p, Zm-q values of GCNF (Zm-o horizontal line, Zm-p solid line, Zm-q dashed line).

## Competing interests

The authors declare that they have no competing interests.

## Authors' contributions

WF and KW designed the study. WF, ZD and BH designed and performed the computational modeling and drafted the manuscript. All the authors read and approved the final manuscript.

## Acknowledgments

We thank Yunlong Liu on discussion of this research. This work was funded by grants from the National Natural Science Foundation of China (61071174) and from Fundamental Research Funds for the Central Universities (HEUCFT1102).

Received: 18 May 2012 Accepted: 28 September 2012  
Published: 5 October 2012

## References

1. Robertson KD: DNA methylation and human disease. *Nat Rev Genet* 2005, **6**(8):597–610.
2. Guerrero-Bosagna C, Skinner MK: Environmentally induced epigenetic transgenerational inheritance of phenotype and disease. *Mol Cell Endocrinol* 2012, **354**(1–2):3–8.
3. Banelli B, Di Vinci A, Gelvi I, Casciano I, Allemanni G, Bonassi S: DNA methylation in neuroblastic tumors. *Cancer Lett* 2005, **228**(1–2):37–41.
4. Bollati V, Baccarelli A: Environmental epigenetics. *Heredity* 2010, **105**(1):105–112.
5. Davies PC: The epigenome and top-down causation. *Interface Focus* 2012, **2**(1):42–48.
6. Costenbader KH, Gay S, Riquelme ME, Iaccarino L, Doria A: Genes, epigenetic regulation and environmental factors: which is the most relevant in developing autoimmune diseases? *Autoimmun Rev*, in press.

7. Yamagata Y, Szabó P, Szüts D, Bacquet C, Arányi T, Páldi A: Rapid turnover of DNA methylation in human cells. *Epigenetics* 2012, **7**(2):141–145.
8. Zilberman D, Gehring M, Tran RK, Ballinger T, Henikoff S: Genome-wide analysis of *Arabidopsis thaliana* DNA methylation uncovers an interdependence between methylation and transcription. *Nat Genet* 2007, **39**(1):61–69.
9. Palacios D, Summerbell D, Rigby PW, Boyes J: Interplay between DNA methylation and transcription factor availability: implications for developmental activation of the mouse myogenin gene. *Mol Cell Biol* 2010, **30**(15):3805–3815.
10. Lienert F, Wirbelauer C, Som I, Dean A, Mohn F, Schübeler D: Identification of genetic elements that autonomously determine DNA methylation states. *Nat Genet* 2011, **43**(11):1091–1097.
11. Choy MK, Movassagh M, Goh HG, Bennett MR, Down TA, Foo RS: Genome-wide conserved consensus transcription factor binding motifs are hyper-methylated. *BMC Genomics* 2010, **11**:519–528.
12. Jones PA, Takai D: The role of DNA methylation in mammalian epigenetics. *Science* 2001, **293**(5532):1068–1070.
13. Kapoor A, Agius F, Zhu JK: Preventing transcriptional gene silencing by active DNA demethylation. *FEBS Lett* 2005, **579**(26):5889–5898.
14. Hertzberg L, Izraeli S, Domany E: STOP: searching for transcription factor motifs using gene expression. *Bioinformatics* 2007, **23**(14):1737–1743.
15. Gibney ER, Nolan CM: Epigenetics and gene expression. *Heredity* 2010, **105**:4–13.
16. Campanero MR, Armstrong MI, Flemington EK: CpG methylation as a mechanism for the regulation of E2F activity. *PNAS* 2000, **97**(12):6481–6486.
17. Comb M, Goodman HM: CpG methylation inhibits proenkephalin gene expression and binding of the transcription factor AP-2. *Nucleic Acids Res* 1990, **18**(13):3975–3982.
18. Mancini DN, Singh SM, Archer TK, Rodenhiser DI: Site-specific DNA methylation in the neurofibromatosis (NF1) promoter interferes with binding of CREB and SP1 transcription factors. *Oncogene* 1999, **18**(28):4108–4119.
19. Radtke F, Hug M, Georgiev O, Matsuo K, Schaffner: Differential sensitivity of zinc finger transcription factors MTF-1, Sp1 and Krox-20 to CpG methylation of their binding sites. *Biol Chem Hoppe Seyler* 1996, **377**(1):47–56.
20. Fang F, Fan S, Zhang X, Zhang MQ: Predicting methylation status of CpG islands in the human brain. *Bioinformatics* 2006, **22**(18):2204–2209.
21. Yin H, Blanchard KL: DNA methylation represses the expression of the human erythropoietin gene by two different mechanisms. *Blood* 2000, **95**(1):111–119.
22. Gu P, Le Menuet D, Chung AC, Cooney AJ: Differential recruitment of methylated CpG binding domains by the orphan receptor GCNF initiates the repression and silencing of Oct4 expression. *Mol Cell Biol* 2006, **26**(24):9471–9483.
23. Zechel C: The Germ Cell Nuclear Factor (GCNF). *Mol Reprod Dev* 2005, **72**:550–556.
24. Thomas AD, Vardhman KR, Daniel JT, et al: A Bayesian deconvolution strategy for immunoprecipitation-based DNA methylome analysis. *Nat Biotechnol* 2008, **7**:779–785.
25. Kim K, Oh M, Ki H, Wang T, Bareiss S, Fini ME: Identification of E2F1 as a positive transcriptional regulator for  $\delta$ -catenin. *Biochem Biophys Res Commun* 2008, **369**(2):414–420.
26. Bozek K, Relógio A, Kielbasa SM, Heine M, Dame C, Kramer A: Regulation of clock-controlled genes in mammals. *PLoS One* 2009, **4**(3):e4882.

doi:10.1186/1471-2164-13-532

**Cite this article as:** Feng et al.: Analysis method of epigenetic DNA methylation to dynamically investigate the functional activity of transcription factors in gene expression. *BMC Genomics* 2012 **13**:532.