

Kansas State University Libraries

New Prairie Press

Conference on Applied Statistics in Agriculture

2008 - 20th Annual Conference Proceedings

A TWO-STAGE APPROACH FOR ESTIMATING THE EFFECT OF DNA METHYLATION ON DIFFERENTIAL EXPRESSION USING TILING ARRAY TECHNOLOGY

Suk-Young Yoo

R. W. Doerge

Follow this and additional works at: <https://newprairiepress.org/agstatconference>



Part of the [Agriculture Commons](#), and the [Applied Statistics Commons](#)



This work is licensed under a [Creative Commons Attribution-Noncommercial-No Derivative Works 4.0 License](#).

Recommended Citation

Yoo, Suk-Young and W., R. Doerge (2008). "A TWO-STAGE APPROACH FOR ESTIMATING THE EFFECT OF DNA METHYLATION ON DIFFERENTIAL EXPRESSION USING TILING ARRAY TECHNOLOGY," *Conference on Applied Statistics in Agriculture*. <https://doi.org/10.4148/2475-7772.1101>

This is brought to you for free and open access by the Conferences at New Prairie Press. It has been accepted for inclusion in Conference on Applied Statistics in Agriculture by an authorized administrator of New Prairie Press. For more information, please contact cads@k-state.edu.

A TWO-STAGE APPROACH FOR ESTIMATING THE EFFECT OF DNA METHYLATION ON DIFFERENTIAL EXPRESSION USING TILING ARRAY TECHNOLOGY

Suk-Young Yoo and R. W. Doerge
Department of Statistics
Purdue University
150 North University Street
West Lafayette, IN 47907 USA

Abstract

Epigenetics is the study of heritable alterations in gene function without changing the DNA sequence itself. It is known that epigenetic modifications such as DNA methylation and histone modifications are highly correlated with the regulation of gene expression. A two-stage analysis is proposed that employs a hidden Markov model and a linear model to evaluate differential expression as related to DNA methylation for the purpose of examining the effects of DNA methylation on gene regulation using tiling array technology. In the first stage, a hidden Markov model (HMM) is employed to estimate the methylation status per tile by utilizing information of neighboring tiles. In the second stage, a linear model is applied to the expression data to identify significantly differentially expressed tiles given the changes in methylation status. The two-stage analysis is applied to *Arabidopsis* chromosome 4 tiling array data and the results are compared with a traditional ANOVA model which has been employed to identify significantly differentially expressed genes using microarray data.

Keywords: epigenetics, epigenomics, methylation, histone modification, tiling array, hidden Markov model, ANOVA model

1. Introduction

Recently, there has been growing interest in investigating epigenetic mechanisms as related to regulation of gene expression. Epigenetics is the study of heritable alterations in gene function without changes in nucleotide sequence. Epigenomics is the genome-wide study of the distribution of epigenetic modifications in the genome. DNA methylation and histone modifications are the two main examples of epigenetic modifications and it is known that epigenetic modifications are highly correlated with regulation of gene expression, genome imprinting and gene silencing (Vaillant and Paszkowski 2007). DNA methylation is achieved by attaching a methyl group (CH_3) to DNA, typically at cytosine bases (C) (Figure 1). In plants, DNA heterochromatin, which has been defined as a tightly packed chromosomal region (Heitz 1928), is heavily methylated and is associated with gene silencing (Lippman et

al. 2004). It has been demonstrated that cytosine methylation controls gene expression of developing seeds, flowering time and floral morphogenesis (Hakabatashi et al. 2005). Similarly, histone modifications occur when a different combination of molecules (i.e., methyl (Me), acetyl (Ac3), etc.) attach to a histone tail (Figure 1). Histone modifications also play an important role in transcriptional stimulation or repression. For example, in plants, dimethylation of lysine 9 and lysine 27 of histone H3 (i.e., H3mK9, H3mK27) with heavy DNA methylation is associated with the transcription repression in heterochromatin. On the other hand, dimethylation of lysine 4 of histone H3 (i.e., H3mK4) is connected with transcription activation in euchromatin (a lightly packed chromosomal region) (Grewal and Jia 2007; Kouzarides 2007; Vaillant et al. 2007). Furthermore, it has been shown that histone acetylation and deacetylation correlate with gene activation and repression, respectively (Nelissen et al. 2007). In this study, we focus on investigating the correlation between DNA methylation and gene expression.

Microarray technology has been widely used to monitor tens of thousands of genes simultaneously. It has also been successfully employed in epigenomic research for evaluating epigenomic modifications. To detect methylated regions of the genome a sample is split into two subsamples. One subsample acts as the control (i.e., methylation in its unaltered state) and the other subsample is treated. Typical treatments for establishing methylation status are either McrBC (methylation-dependent restriction enzyme) or bisulfite modification. For the bisulfite modification treatment, unmethylated cytosines (C) are converted to uracils (U), whereas McrBC cleaves DNA containing methylated C's (Shi et al. 2003; Lippman et al. 2005). In order to identify which features are methylated, McrBC or bisulfite treated and untreated DNA samples are hybridized to microarrays and by comparing treated and untreated samples, methylation status can be evaluated.

Tiling microarray technology (Mockler and Ecker 2004) has become a popular tool for studying gene expression, methylation, and histone modifications across a whole genome with high-density probes (Emanuelsson et al. 2007; Gendrel et al. 2005; Johnson et al. 2006; Li et al. 2008; Lippman et al. 2005; Schumacher et al. 2006; Wu et al., 2006; Zhang et al. 2006; Zilberman et al. 2007). Tiling arrays are similar to cDNA and oligo microarrays, but consist of DNA sequences or probes that are not necessarily genes. Unlike cDNA or oligo arrays, tiling arrays are comprised of linearly ordered and sometimes overlapping tiles. Because of the underlying linear order, neighboring tiles can be correlated with each other. Effectively utilizing information from neighboring tiles is one of the current statistical challenges when studying biological processes using tiling array technology.

Although epigenomic research has added to our understanding of biological phenomena, to date little work has been done to establish novel statistical methods that exploit the relationship between epigenetic modifications and gene expression. In this research, we concentrate on differential expression (i.e., transcription) and DNA methylation

tiling array experiments to introduce an ANOVA model and hypothesis tests for analyzing the individual tiling array data that result from each unique investigation (i.e., expression and methylation). A novel statistical approach, a two-stage analysis, is proposed to investigate the effects of methylation status on gene regulation. Our two-stage analysis employs a hidden Markov model and a linear model to combine data and results from both methylation and differential expression tiling arrays.

2. Analysis of tiling array data

Differential Expression: Analysis of variance (ANOVA) models (Kerr et al., 2000; Kerr and Churchill, 2001) have been employed to analyze tiling microarray data. When dye-labeled, an experimental design (e.g., dye-swap) that is popular with cDNA microarray analysis can be employed. The ANOVA model using a common variance assumption (i.e., all tiles have the same level of variation) is:

$$\log(y_{ijkgr}) = \mu + A_i + D_j + G_k + T_g + AT_{ig} + DT_{jg} + GT_{kg} + \varepsilon_{ijkgr} \quad (2.1)$$

where $i = 1, \dots, n_a$; $j = 1, 2$; $k = 1, 2$; $g = 1, \dots, n_{ug}$, and $r = 1, \dots, n_g$; μ is the grand mean, and A, D, G and T are the array, dye, genotype (or treatment) and tile effects, respectively. Moreover, AT, DT and GT are the interactions between array and tile, dye and tile, and treatment and tile, respectively. The number of arrays is denoted as n_a , n_{ug} is the number of unique tiles, and n_g is number of replicates of each tile. The error term, ε_{ijkgr} , is an independent random variable from a normal distribution with mean 0 and variance σ^2 . The hypotheses tested for identifying differentially expressed (i.e., transcription) tiles are:

$$H_0 : G_1 + GT_{1g} = G_2 + GT_{2g} \quad \text{vs} \quad H_1 : G_1 + GT_{1g} \neq G_2 + GT_{2g} \quad (2.2)$$

where G and GT are treatment effect and interaction between treatment and tile (Black, 2002).

Methylation: To identify significantly methylated tiles, samples treated with McrBC or bisulfite are compared with untreated samples (Shi et al., 2003; Lippman et al., 2005; Schumacher et al., 2006). To identify significantly methylated tiles, control tiles are often used for relative comparisons. Control tiles are selected based on their lack of methylation (confirmed by independent laboratory techniques) and the median difference of intensities between treated and untreated control tiles serves as a basis for testing methylation. The hypotheses for identifying significantly differentially methylated tiles, adjusted using control tiles, are:

$$H_0 : (G_1 + GT_{1g}) - (G_2 + GT_{2g}) = \mu_0 \quad \text{vs} \quad H_1 : (G_1 + GT_{1g}) - (G_2 + GT_{2g}) > \mu_0 \quad (2.3)$$

where $\mu_0 = \text{median}\{(G_1 + GT_{1g}) - (G_2 + GT_{2g}), g \text{ is a control}\}$. (2.4)

A one-tailed t-test is performed to test the hypotheses (2.3) since methylation can be detected or not (i.e., one direction). This is in contrast to differential expression tiling array experiments where tiles can be down- or up-regulated (i.e., both directions) when two genotypes or samples are compared, and thus a two-tailed test is employed.

3. A two-stage analysis

3.1 First stage: a hidden Markov model

In the first stage of the proposed two-stage analysis, a hidden Markov model (HMM) is employed to estimate the methylation status of each tile. Since tiles span the genome in an ordered linear fashion, information from neighboring tiles may add valuable information to the evaluation of per tile expression or methylation. In the HMM, the methylation status is defined as hidden states and information from adjacent tiles is incorporated by including their methylation status in the hidden states. Based on this, hidden states are combinations of methylation status of three tiles (i.e., tiles, $i-1, i, i+1$; Figure 2). The observable output is the observed methylation status of tile i . Using defined hidden states, the methylation status of the set of adjacent tiles (tiles $i-1, i+1$), as well as the methylation status of tile i , have an effect when methylation status is observed for tile i . Since the methylation status of each tile is a binary variable (i.e., 1 for methylated tiles and 0 for unmethylated tiles), there are 8 possible hidden states. The hidden states employed in this application are illustrated in Figure 2. Based on the hidden states in Figure 2, the parameters of HMM can be defined (Rabiner 1989). Let N be the number of hidden states ($N = 8$). Let M denote the number of distinct observations per state. Since each tile can be 0 if the tile i is not methylated and 1 if tile i is methylated, $M=2$. Define $S = \{S_1, \dots, S_8\}$ to be the individual hidden states. Specifically,

$$S = \{S_1, \dots, S_8\} = \{(1,1,1), (1,1,0), (1,0,1), (0,1,1), (1,0,0), (0,1,0), (0,0,1), (0,0,0)\}.$$

Let $V = \{v_1, v_2\}$ denote the observed methylation status of tile i and $V = \{0, 1\}$. If q_t is the state at time t , then the initial state distribution is

$$\pi = P\{q_1 = S_i\}, 1 \leq i \leq 8. \quad (3.1)$$

The state transition probability distribution is

$$a_{ij} = P[q_{t+1} = S_j | q_t = S_i], 1 \leq i, j \leq 8, \quad \text{and} \quad (3.2)$$

the observation probability distribution in state j is

$$b_j = P[v_k \text{ at } t | q_t = S_j], 1 \leq j \leq 8, k = 0,1. \quad (3.3)$$

Using the Viterbi algorithm, the methylation status is estimated for the given observed methylation status of tiles (Viterbi 1967).

3.2 Second stage: an ANOVA model

The second stage of the proposed approach incorporates the estimated methylation status for each tile (as was determined in the first stage) into an ANOVA model for the purpose of testing each tile for differential expression given the changes in methylation status. The ANOVA model for the second stage is:

$$\log(y_{ijkl}) = \mu + X_i + M_j + T_k + XM_{ij} + XT_{ik} + MT_{jk} + \varepsilon_{ijkl} \quad (3.4)$$

where $i = 0,1$; $j = 0,1$; $k = 1, \dots, n$, and $l = 1, \dots, a$; μ is the grand mean, and X, M, and T are the genotype, methylation, and tile effects, respectively. Additionally, XM, XT and MT are the interactions between treatment and methylation, treatment and tile, and methylation and tile, respectively. The number of unique tiles is denoted by n and a is the number of replicates per tile. ε_{ijkl} is the error term which is an independent random variable from a normal

distribution with mean 0 and variance σ^2 . The hypotheses for identifying significantly differentially expressed tiles given the changes in methylation status are

$$H_0 : M_0 + MT_{1k} = M_1 + MT_{1k} \text{ vs } H_1 : M_0 + MT_{0k} \neq M_1 + MT_{1k}. \quad (3.5)$$

To accommodate the multiple testing issues, the false discovery rate (FDR) and Holm procedures are used to control the significance level, $\alpha = 0.05$ (Holm 1979; Benjamini and Hochberg 1995).

3.3 Application to *Arabidopsis* chromosome 4 tiling array data

The proposed two-stage analysis was implemented in the software package R and applied to the *Arabidopsis* chromosome 4 tiling array data from Lippman et al. (2004). In this study, tiling array experiments were conducted to test for differential expression between wild type and *ddm1* (decreased in DNA methylation 1) mutant, as well as test for methylation within both wild type and *ddm1*. The same tiling array design was employed for all three studies (i.e., differential expression, methylation status for wild type, and methylation status for *ddm1*). In other words, the same tiles were employed to study both differential expression and DNA methylation. There were 1724 unique tiles which spanned chromosome 4 in an ordered, overlapping linear manner. For each experiment, tiles were replicated two to six times so that the total number of tiles in each experiment is 3822 or 5724. For differential expression, 6 arrays were employed (one and two dye swaps using two biological replicates), while 4 arrays were used in the methylation tiling array experiments (two dye

swaps using one biological sample). First, an ANOVA model (2.1) was performed for the individual analyses of differential expression and methylation. After FDR adjustment, 785 tiles were identified as significantly differentially expressed between wild type and *ddm1*. For DNA methylation, McrBC treated and untreated DNA samples were compared for both wild type and *ddm1* independently (i.e., two separate experiment). 800 tiles in the euchromatin were employed as control tiles, and the median difference of intensity (as defined in equation (2.4)) between McrBC treated and untreated control tiles was calculated for the hypothesis tests (2.3). In wild type 743 tiles were significantly methylated by adjusting FDR, whereas only 152 tiles were found methylated in *ddm1*, which are predictable results since DNA methylation is depleted in *ddm1*.

In order to understand the value of the proposed two-stage analysis it was employed to test differential expression given the methylation status of wild type and *ddm1*. Methylation status was estimated using a HMM (parameters of HMM are shown in Table 1). The methylation results from the first stage were include in the ANOVA model (3.4) to identify statistically significantly differentially expressed tiles as related to methylation status. Using an FDR adjustment 286 tiles were found statistically significantly differentially expressed given their methylation status. The results from the proposed two-stage analysis are compared with the results from a traditional ANOVA analysis (2.1) for differential expression (Table 2 and Figure 3). It is known that the region identified between tile IDs 401 and 1000 on chromosome 4 is a heterochromatic region. The statistically significant differentially expressed tiles (273 using Holm and 785 using FDR) from an ANOVA (2.1) are distributed across the whole chromosome 4 (Figure 3a). In contrast, the majority of the statistically significant tiles (241 out of 286 using FDR) identified from the two-stage analyses are located in the heterochromatin region (Figure 3b). Specifically in wild type, tiles are heavily methylated in the heterochromatic region, but in *ddm1* methylation is significantly depleted for many tiles. Thus, changes in methylation status were detected for many tiles found in the heterochromatin. These results provide evidence that the presence of DNA methylation is highly correlated with repression of transcription.

4. Summary

A novel statistical method was developed to integrate the results from methylation with differential expression data using tiling array technology. Traditionally, when investigating the relationship between differential expression and epigenetic modifications (i.e., DNA methylation and histone modifications), individual analyses are conducted and results combined graphically to provide visual summaries (Lippman et al. 2004; Zhang et al. 2006; Li et al. 2008). Through this research a novel two-stage analysis combined results from DNA methylation with data from differential expression tiling array experiment using a linear statistical model (ANOVA) to test the effect of DNA methylation on regulation of gene

expression. By employing a traditional ANOVA model (2.1) significantly differentially expressed tiles between two samples (here, genotypes) can be found. On the other hand, by applying the proposed two-stage analysis, it is possible to identify tiles that are significantly differentially expressed given their methylation status. Other work that has been completed and presented elsewhere, or that is currently underway includes simulations that explore both the statistical power of the two-stage analysis, as well as the benefit of incorporating histone modifications into the second stage of the proposed approach.

Acknowledgments

We thank Professor Rob Martienssen and his laboratory at Cold Spring Harbor Laboratory for providing the *Arabidopsis* chromosome 4 tiling array data.

References

Emanuelsson, O., Nagalakshmi U., Zheng D., Rozowsky, J. S., Urban, A. E., Du, J., Lian, Z., Stolc, V., Weissman, S., Snyder, M. and Gerstein, M. B., 2007. Assessing the performance of different high-density tiling microarray strategies for mapping transcribed regions of the human genome. *Genome Research* 17:886-897.

Gendrel, A.-V., Lippman, Z., Martienssen, R. and Colot, V., 2005. Profiling histone modification patterns in plants using genomic tiling microarrays. *Nature methods* 2:213-218.

Grewal, S. I., and Jia, S., 2007. Heterochromatin revisited. *Nature Reviews Genetics* 8:35-46.

Hakabayashi, K., Okamoto, M., Koshiba, T., Kamiya, Y. and Nambara, E., 2005. Genome-wide profiling of stored mRNA in *Arabidopsis thaliana* seed germination: epigenetic and genetic regulation of transcription in seed. *The Plant Journal* 41(5):697-709.

Heitz, E., 1928. Das heterochromatin der Moose. *Jehrb. Wiss. Botanik* 69,:762–818.

Holm, S., 1979 A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics* 6:65-70.

Johnson, W. E., Li, W., Meyer, C. A., Gottardo, R., Carroll, J. S., Brown, M. and Liu, X. S., 2006. Model-based analysis of tiling-arrays for ChIP-chip. *PNAS* 103(33):12457-12462.

Kerr, M.K., Martin, M and Churchill, G. A., 2000. Analysis of variance for gene expression microarray data. *Journal of Computational Biology* 7:819-937.

Kerr, M.K., and Churchill, G. A., 2001. Statistical design and the analysis of gene expression microarray data. *Genetical Research* 77:123-128.

Kouzarides, T., 2007 Chromatin Modifications and their Function. *Cell* 128: 693-705.

Li, X., Wang, X., He, K., Ma, Y., Su, N., Hang, H., Stole, V., Tongprasit, W., Jin, W., Jiang, J., Terzaghi, W., Li, S., and Deng, X. W., 2008. High-resolution mapping of epigenetic modification of the rice genome uncovers interplay between DNA methylation, histone modification, and gene expression. *The Plant Cell online publication*.

Lippman, Z., Gendrel, A-V., Black, M., Vaughn, M. W., Dedhia, N., McCombie, W. R., Lavine, K., Mittal, V., May, B., Kasschau, K. D., Carrington, J. C., Doerge, R. W., Colot, V. and Martienssen, R., 2004. Role of transposable elements in heterochromatin and epigenetic control. *Nature* 430:471-476.

Mockler, T. C. and Ecker, J. R., 2005. Applications of DNA tiling arrays for whole-genome analysis. *Genomics* 85(1):1-15.

Nelissen, H., Boccardi, T. M. and Himanen, K., 2007. Impact of Core Histone Modifications on Transcriptional Regulation and Plant Growth. *Critical Reviews in Plant Sciences* 26:243–263.

Rabiner, L. R., 1989. A Tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE* 77(2):257-286

Schumacher, A., Kapranov, P., Kaminsky, Z., Flanagan, J., Assadzadeli, A., Yau, P., Virtanen, C., Winegarden, N., Cheng, J., Gingeras, T. and Petronis, A., 2006. Microarray-based DNA methylation profiling: technology and applications. *Nucleic Acids Research* 34(2):528-542.

Shi, H., Maier, S., Nimmrich, I., Yan, P. S., Caldwell, C. W., Olek, A., and Huang, T. H.-M., 2003. Oligonucleotide-based microarray for DNA methylation analysis: principles and applications. *Journal of Cellular Biochemistry* 88:138-143.

Vaillant, I., and Paszkowski, J., 2007 Role of histone and DNA methylation in gene regulation. *Current Opinion in Plant Biology* 10:528-533.

Viterbi, A. J., 1967. Error bounds for convolution codes and an asymptotically optimum decoding algorithm. *IEEE Transaction on Information Theory* 13:260-269.

Wu, J., Smith, L. T., Plass, C. and Huang, T. H-M., 2006. ChIP-chip comes of age for genome-wide functional analysis. *Cancer Research* 66(14):6899-6902.

Zhang, X., Yazaki, J., Sundaresan, A., Cokus, S., Chan, S. W.-L., Chen, H., Henderson, I. R., Shinn, P., Pellegrini, M., Jacobsen, S. E., and Ecker, J. R., 2006. Genome-wide high-resolution mapping and functional analysis of DNA methylation in *Arabidopsis*. *Cell* 126:1189-1201.

Zilberman, D., Gehring, M., Robert, T. K., Ballinger, T. and Henikoff, S., 2007. Genome-wide analysis of *Arabidopsis thaliana* DNA methylation uncover an interdependence between methylation and transcription. *Nature Genetics* 39(1):61-69.

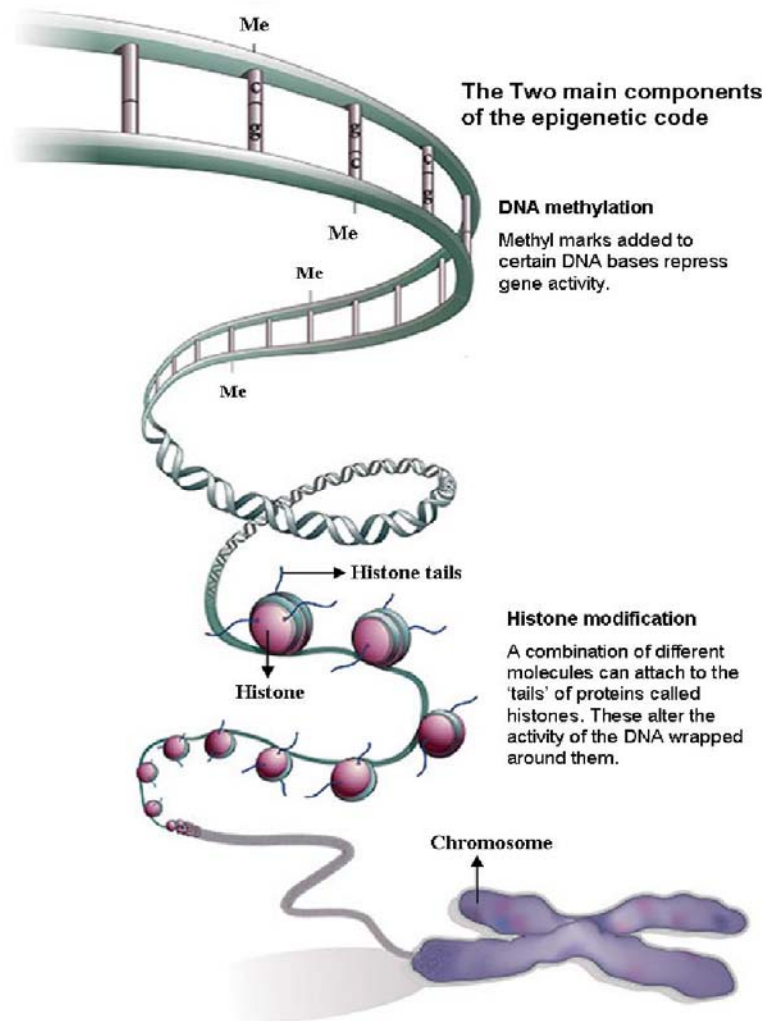


Figure 1. Two examples of epigenetic modifications. DNA methylation occurs by adding a methyl group (CH_3) to a DNA base, typically at base C. Generally, DNA methylation represses transcription. Histone modifications occur by attaching various molecules (e.g., methyl and acetyl) to histone tails which in turn may influence the activity of the DNA wrapped around the histone (Shi et al. 2003).

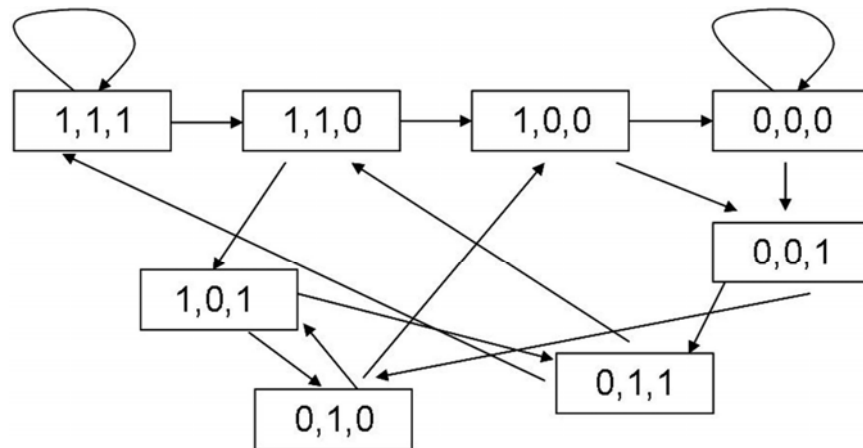


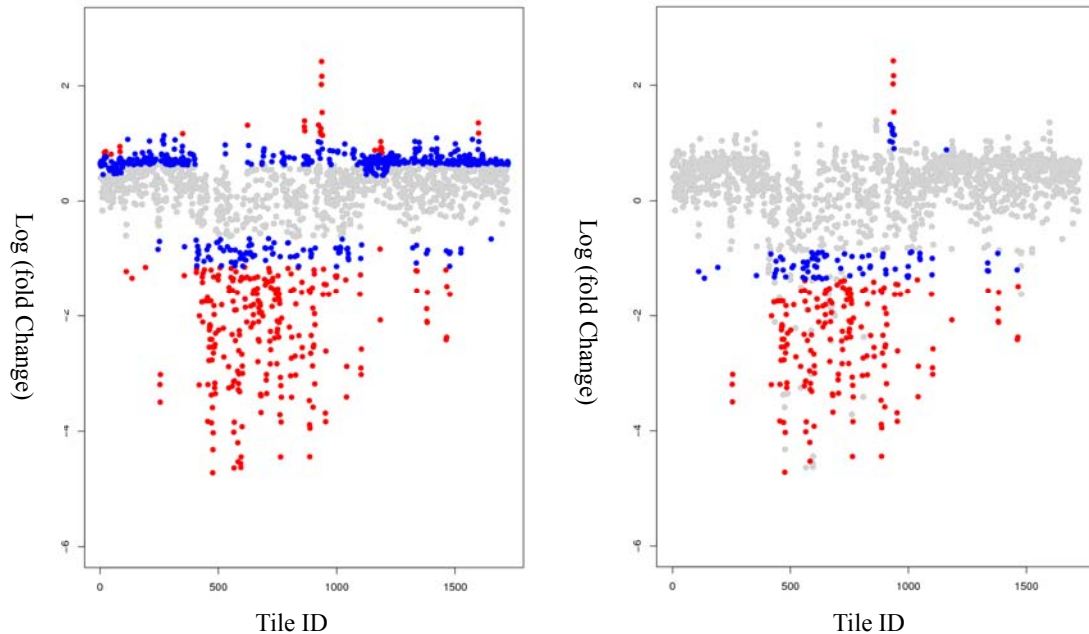
Figure 2. All possible hidden states for a hidden Markov model (HMM). Each state represents the methylation status of tiles $i-1$, i , $i+1$ in single base pair transitions from left to right through the possible methylated DNA base pairs (i.e., the $i+1$ state in the current state becomes the i^{th} state in the transition state). A binary variable represents the methylation status of each tile, where 0 is an unmethylated tile and 1 denotes a methylated tile. Arrows represent all possible transitions from state to state.

Table 1. Parameters used for the HMM in the first stage of the proposed two-stage analysis. The initial state distribution, the observation probability distribution, and the transition probability distribution used for the HMM in the two-stage analysis of *Arabidopsis* chromosome 4 tiling array data are presented.

Initial state distribution	Observation probability distribution	
$P((1,1,1))=0.1557$	$P(0 (1,1,1))=0.1$	$P(1 (1,1,1))=0.9$
$P((1,1,0))=0.0374$	$P(0 (1,1,0))=0.3$	$P(1 (1,1,0))=0.7$
$P((1,0,1))=0.0113$	$P(0 (1,0,1))=0.5$	$P(1 (1,0,1))=0.5$
$P((0,1,1))=0.0374$	$P(0 (0,1,1))=0.3$	$P(1 (0,1,1))=0.7$
$P((1,0,0))=0.0551$	$P(0 (1,0,0))=0.7$	$P(1 (1,0,0))=0.3$
$P((0,1,0))=0.029$	$P(0 (0,1,0))=0.4$	$P(1 (0,1,0))=0.6$
$P((0,0,1))=0.0551$	$P(0 (0,0,1))=0.7$	$P(1 (0,0,1))=0.3$
$P((0,0,0))=0.6189$	$P(0 (0,0,0))=0.95$	$P(1 (0,0,0))=0.05$
Transition probability matrix		
$\begin{bmatrix} 0.86 & 0.14 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.2 & 0 & 0.8 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.64 & 0 & 0.36 & 0 & 0 \\ 0.58 & 0.42 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.16 & 0.84 \\ 0 & 0 & 0.14 & 0 & 0.86 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.55 & 0 & 0.45 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.08 & 0.92 \end{bmatrix}$		

Table 2. Results from a traditional ANOVA model (2.1) and the two-stage analysis (3.4). Number of significant tiles identified using Holm and FDR adjustments are presented. The numbers in parentheses are the number of significant tiles located in the heterochromatin.

	ANOVA model (2.1)	ANOVA model (3.4)
Holm	273 (229)	173 (154)
FDR	785 (329)	286 (241)



(a) Results from a traditional ANOVA (2.1) (b) Results from a two-stage analysis

Figure 3. Results from the traditional ANOVA model (2.1) are shown in (a). Results from the ANOVA model (3.4) are shown in (b). The data are *Arabidopsis* chromosome 4 tiling array data (Lippman et al. 2004). In both plots, the Y axis is an average of log fold changes for each tile and the X axis is tile ID. Moreover, red points represent significantly identified tiles using both Holm and FDR and blue points represent significantly identified tiles using FDR only. Positive values of log fold changes indicate that tiles are more expressed in wild type when compared to *dml1*. The region between tiles ID between 401 and 1000 is the heterochromatin region. (a): Significant tiles identified using an ANOVA model (2.1) are distributed across the whole region of chromosome 4. Most significant tiles in the heterochromatin region are downregulated. (b): The majority of the significant tiles are located in the heterochromatin region, and those tiles are significantly differentially expressed given changes in methylation status.