

Article

Detection of Differentially Methylated Regions Using Bayes Factor for Ordinal Group Responses

Fengjiao Dunbar ¹, Hongyan Xu ², Duchwan Ryu ³, Santu Ghosh ², Huidong Shi ⁴
and Varghese George ^{2,*}

¹ Genomics Research Center, AbbVie, North Chicago, IL 60064, USA; fengjiao.dunbar@abbvie.com

² Department of Population Health Sciences, Augusta University, Augusta, GA 30912, USA; hxu@augusta.edu (H.X.); sghosh@augusta.edu (S.G.)

³ Department of Statistics and Actuarial Science, Northern Illinois University, DeKalb, IL 60178, USA; dryu@niu.edu

⁴ Georgia Cancer Center, Augusta University, Augusta, GA 30912, USA; hshi@augusta.edu

* Correspondence: vgeorge@augusta.edu; Tel.: +1-706-721-0801

Received: 19 July 2019; Accepted: 15 September 2019; Published: 17 September 2019

Abstract: Researchers in genomics are increasingly interested in epigenetic factors such as DNA methylation, because they play an important role in regulating gene expression without changes in the DNA sequence. There have been significant advances in developing statistical methods to detect differentially methylated regions (DMRs) associated with binary disease status. Most of these methods are being developed for detecting differential methylation rates between cases and controls. We consider multiple severity levels of disease, and develop a Bayesian statistical method to detect the region with increasing (or decreasing) methylation rates as the disease severity increases. Patients are classified into more than two groups, based on the disease severity (e.g., stages of cancer), and DMRs are detected by using moving windows along the genome. Within each window, the Bayes factor is calculated to test the hypothesis of monotonic increase in methylation rates corresponding to severity of the disease versus no difference. A mixed-effect model is used to incorporate the correlation of methylation rates of nearby CpG sites in the region. Results from extensive simulation indicate that our proposed method is statistically valid and reasonably powerful. We demonstrate our approach on a bisulfite sequencing dataset from a chronic lymphocytic leukemia (CLL) study.

Keywords: Bayes factor; Bayesian mixed-effect model; CpG sites; DNA methylation; Ordinal responses

1. Introduction

It is now widely accepted that cancer develops through a series of stages [1]. It starts from a very limited area, not invasive and metastatic at the early stage, then spreads to distant sites in the body, and becomes highly invasive and metastatic at the late stage. In addition, patient survival times are significantly reduced at the late stages. For example, the 5-year relative survival rate for lung cancer is 54% at a localized stage, and is reduced to 4% at the distant stage [2]. More than half of lung cancers are diagnosed at a distant stage, which indicates that early diagnosis of cancer is the main factor to enhance patient survival. Therefore, markers for early detection and proper classification of the tumor are extremely critical to improve life expectancy. Furthermore, identifying high-risk cancer patients at an early stage, would allow them to receive standard chemotherapy in advance.

DNA methylation has been found to be a marker for disease diagnosis, such as in cancer [3]. Significant progress has been made using DNA methylation differences to capture substantial

information about the molecular and gene-regulatory states among biology subtypes, such as tumor and normal tissues [4].

In addition, DNA methylation can be used as a marker to differentiate disease severity, such as early and late stages in breast cancer [5], ovarian cancer [6] and prostate cancer [7]. Most of them have potential functions in inducing and suppressing cancer metastasis. Moreover, DNA methylation is associated with tumor size in colorectal cancer [8]. Patients with higher methylation showed more frequent recurrence as compared with the low-methylation group, and shortened cancer-related survival and recurrence-free survival [8].

These findings show the critical importance of a better understanding of cancer progression and metastasis, which could help make better prediction of the clinical aggressiveness of cancer. Since DNA methylation is associated with disease severity, detecting differentially methylated regions (DMRs) can help understand cancer progression.

Most analyses are conducted by creating dichotomies based on biological subtypes, such as early and late cancer stages, and then detect DMRs by comparing the differences of DNA methylation rates between two groups [5–7]. However, when there are actually more than two groups, such approaches may lose information regarding multiple disease status, due to collapsing or ignoring clinically relevant subtypes, resulting in suboptimal clinical conclusions and decisions.

To use multiple disease status, it is possible to run multiple testing for the association between DNA methylation and multiple group responses, using the methods for two groups. Although we can simply run analysis for all pair-wise comparisons and combine the results, it is not trivial when considering the regional correlation of DMRs, and would increase the multiple testing burden.

Another possible method is the generalized linear model that includes indicator variables for different levels of disease status. This method has the advantage that it can adjust for covariates. However analysts are often faced with noisy estimates of category-specific regression coefficients, which can lead to unreasonable patterns in the regression coefficients corresponding to different levels of disease status, and it can reduce the power [9].

To improve the efficacy of an overall test, one can take advantage of the fact that cancer develops through a series of stages, or different levels of disease severity in general, and develop statistical methods that can incorporate the ordering of disease status. However, the widely used trend test is not an ideal method, because it requires scores or weights for different levels of disease status, which are generally unknown.

Here we propose a Bayesian approach and use the Bayes factor to test the association between methylation rates and disease severity. The proposed Bayes Factor Method (BFM) can incorporate monotonicity constraints, and find DMRs in which methylation rates increase (or decrease) as the diseases become more severe. Patients are classified into groups based on the disease severity (e.g., stages of cancer), and DMRs are detected by using moving windows along the genome. Within each window, the Bayes factor is calculated and is used to test the hypothesis of constant versus monotonic increase in methylation rates corresponding to the severity of the disease.

In addition, since DNA methylation rates have been shown to be correlated at nearby CpG sites with complicated correlation structure [10], a linear mixed-effect model is used to incorporate the correlation of methylation rates between and within CpG sites in the region.

2. Materials and Methods

2.1. Methods

Classical statistical inference under constrained parametric spaces has been addressed by many studies. Among them, Bartholomew [11] presented one of the first tests for K multinomial proportions with inequality constraints. He proposed a test of $H_0: p_1 = p_2 = \dots = p_K$ against the simple ordered $H_1: p_1 \leq p_2 \leq \dots \leq p_K$ with at least one strict inequality, where p_k ($k = 1, 2, \dots, K$) represents the proportion the k^{th} group. Under H_0 , the maximum likelihood estimator of p_k is the overall sample proportion π_k . If the sample multinomial proportions satisfy $\pi_1 \leq \pi_2 \leq \dots \leq \pi_K$, then the order-restricted ML estimator is $\hat{p}_k = \pi_k$. However, sometimes the sample proportions may not

satisfy the ordering $\pi_1 \leq \pi_2 \leq \dots \leq \pi_K$; in that case, calculation of the restricted maximum likelihood estimator (RMLE) is subject to arbitrary orderings of the parameters, and it requires specialized algorithms that are not easily generalizable [9].

Robertson and Wegman [12] proposed a likelihood ratio statistic for the inequality-constrained binomial problem, which compares parameters for independent samples from a single-parameter exponential family distribution. Before calculating the test statistic, they used the pool-adjacent-violators algorithm [13] to pool “out-of-order” categories for which $\pi_k > \pi_{k+1}$ until the resulting sample proportions are monotone increasing. The order-restricted ML estimators \hat{p}_k become the adjusted sample proportions.

The idea of applying an isotonic transformation to the unconstrained parameter estimates motivated Dunson and Neelon [9] to create a Bayesian alternative approach for this problem, which has been adapted here. They proposed to use Bayes factors for assessing ordered trends, which are calculated based on the output from Gibbs sampling. The samples from the order-constrained model are derived by transforming samples draws from an unconstrained posterior density using an isotonic regression transformation. Next, we explain our proposed Bayes factor method (BFM).

Suppose m_{kij} is the count of methylated molecules at CpG site j of individual i in group k . We assume $m_{kij} \sim B(c_{kij}, p_{kij})$, where c_{kij} is the coverage, and p_{kij} is the true methylation rate at that particular site, with $k = 1, 2, \dots, K$, $i = 1, 2, \dots, n_k$ and $j = 1, 2, \dots, m$.

Within each moving window along the genome, a mixed-effect model is considered to allow the correlation of methylation rates between and within CpG sites. The logit link function for the methylation rate p_{kij} is expressed by

$$\text{logit}(p_{kij}) = \mu_k + v_{0ki} + v_{1kij}, \quad (1)$$

where v_{0ki} and v_{1kij} are the random effects. The random effect $v_{0ki} \sim N(0, \sigma_{v_0}^2)$ is used to model the interindividual correlation of methylation rates within each CpG site, while the random effect $\mathbf{v}_{1ki} = (v_{1ki1}, v_{1ki2}, \dots, v_{1kim})^T \sim N(\boldsymbol{\mu}_0, \boldsymbol{\Sigma})$, with $\boldsymbol{\mu}_0 = (0, 0, \dots, 0)^T$ is used to model the correlation of methylation rates between CpG sites.

Here μ_k in (1) is the fixed effect for each group, representing the association between methylation rates and group responses. The strength and direction of the association is modeled by prior distribution $N(\mu_\mu, \sigma_\mu^2)$, which means the parameters of μ_μ and σ_μ^2 control the distribution of μ_k , and implies that all of the methylation rates are drawn from a common distribution. This brings the advantage of allowing for heterogeneity of effects across CpG sites, instead of just pooling information across CpG sites in a region. Pooling assumes that each CpG site in the region has same methylation rates, while BFM considers the methylation rates of each CpG sites to be a random quantity governed by a prior distribution.

With assigned hyperpriors $\mu_k \sim N(0, 1000^2)$, $\sigma_k^2 \sim \text{IG}(1, 100)$, $\sigma_{v_0}^2 \sim \text{IG}(1, 100)$ and $\Sigma^{-1} \sim \text{Wish}(I_m, m)$ for m CpG sites in the moving window. The posterior distribution of μ_k is based on the mixed-effect logistic model (1), and it is used to calculate the Bayes factor for comparing the two models, $M_0: \mu_1 = \mu_2 = \dots = \mu_K$, $M_1: \mu_1 \leq \mu_2 \leq \dots \leq \mu_K$ with at least one strict inequality, in order to see whether there is an ordered constraint of methylation rates corresponding to severity of the disease.

To calculate the Bayes factor, first we drew samples $\mu_1, \mu_2, \dots, \mu_K$ from the posterior distribution by using Gibbs sampling. After that, an isotonic transformation is used to transform $\mu_1, \mu_2, \dots, \mu_K$ into $\tilde{\mu}_1, \tilde{\mu}_2, \dots, \tilde{\mu}_K$, with $\tilde{\mu}_1 \leq \tilde{\mu}_2 \leq \dots \leq \tilde{\mu}_K$ [8] by using the min-max formula for the isotonic transformation, given by,

$$\tilde{\mu}_k = g_k(\boldsymbol{\mu}) = \min_{t \in U_k} \max_{s \in L_k} \left(\frac{\mathbf{1}'_{t-s+1} \mathbf{V}_{[s:t]}^{-1} \boldsymbol{\mu}_{[s:t]}}{\mathbf{1}'_{t-s+1} \mathbf{V}_{[s:t]}^{-1} \mathbf{1}_{t-s+1}} \right) \text{ for } j = 1, 2, \dots, K, \quad (2)$$

where $\mathbf{V} = \text{diag} V_1, \dots, V_K$ denotes the posterior covariance matrix and the diagonal submatrix V_i , $i = 1, \dots, k$, is the covariance matrix of the i^{th} ordered group. It is estimated from the samples of the posterior density of $\boldsymbol{\mu}$. U_k and L_k denote subsets of $\{1, \dots, K\}$ such that the ordering $\mu_{j'} \leq \mu_j$ for all $j' \in L_k$ and the ordering $\mu_{j'} \geq \mu_j$ for all $j' \in U_k$. Also samples $\mu_1^0, \mu_2^0, \dots, \mu_K^0$ are drawn from the prior density and transformed into $\tilde{\mu}_1^0, \tilde{\mu}_2^0, \dots, \tilde{\mu}_K^0$, with $\tilde{\mu}_1^0 \leq \tilde{\mu}_2^0 \leq \dots \leq \tilde{\mu}_K^0$, by using the isotonic

transformation in (2). The Bayes factor for each window (with moving windows along the genome) is given by,

$$BF = \frac{P(M_1|data)/P(M_1)}{P(M_0|data)/P(M_0)} = \frac{P(\tilde{\mu}_K > \tilde{\mu}_1)/P(\tilde{\mu}_K^0 > \tilde{\mu}_1^0)}{P(\tilde{\mu}_K = \tilde{\mu}_1)/P(\tilde{\mu}_K^0 = \tilde{\mu}_1^0)}$$

Please note that the isotonic transformation in (2) changes our hypotheses slightly, making the resulting Bayes Factor an approximation rather than exact [14]. The windows with highest value of the Bayes factor among all windows are used for evaluating DMRs.

Thus, the Bayes factor is the ratio of the marginal densities of the data under the two hypotheses, and it can be used to weigh evidence in favor of a hypothesis, by utilizing all the information contained in the full likelihood. Our proposed BFM can detect DMRs associated with disease severity, especially detecting DMRs with monotonically increasing or decreasing methylation rates, as the disease severity increase. It uses a mixed-effect model to not only adjust for correlation of methylation rates between CpG sites within each moving window but also correlations within CpG sites.

In addition, by adding covariates x_{ki} in the model (1), we can account for the effects of covariates that are associated with methylation rates, such as age [15] and gender [16].

To aid in the interpretation of the Bayes factor, Jeffreys [17] proposed the following rule of thumb: “When $3 < BF \leq 10$ the evidence is positive, when $10 < BF \leq 100$ the evidence is strong, and when $BF > 100$, the evidence is decisive”. As Kass and Raftery [18] pointed out, these categories are not precise calibration, but rather a descriptive statement about the standards of evidence in scientific investigations.

2.2. Simulation Study of the Properties of BFM

Extensive simulation was conducted to study the statistical validity and power of BFM to detect DMRs. For simplicity, for each individual, we simulated one CpG island (genomic region with CpG sites) consisting of m equally spaced CpG sites, with only one DMR of length $r (< m)$ in the middle of the island. Further, we used equal sample size, N , for each of the K groups, and, we did not include any covariates.

Simulation Setup:

The goal here is to simulate methylation rate at each CpG site for each individual. This is achieved in two steps. In step 1, methylation data in the form of NGS short reads sequences were simulated for each CpG site, with correlated methylation status between CpG sites. We also assumed that methylation status at CpG sites among different sequences were independent, as expected in NGS data. In step 2, the individual methylation rates were calculated by summarizing the methylation status at each CpG site from the short read sequences.

The simulation details are described below:

First, we generated 100 NGS short reads using 100 pairs of random numbers $\{a, c\}$ where a is the start point and c is the length of each short read sequence.

Then we used vector $\mathbf{Y} = (Y_{kis,a}, Y_{kis,a+1}, \dots, Y_{kis,a+c-1})$ to define the methylation status for short read sequence s of individual i in group k , and generated \mathbf{Y} from a multivariate Bernoulli distribution to allow for the correlation among the methylation rates. $P(\mathbf{Y} = \mathbf{y}) = P(Y_{kisa}, Y_{kisa+1}, \dots, Y_{kisa+c-1})$ of such a discrete random vector \mathbf{Y} depends on 2^c probabilities, $p(0,0, \dots, 0)$, $p(0,0, \dots, 1)$, \dots , $p(1,1, \dots, 1)$, specific to the different realizations of \mathbf{Y} . Considering the fact that if a vector (Y_1, Y_2, \dots, Y_p) follows p -variate Bernoulli distribution, the conditional distribution of (Y_1, Y_2, \dots, Y_r) ($r < p$) given $(Y_{r+1}, Y_{r+2}, \dots, Y_p)$ is also a multivariate Bernoulli distribution [18]. We can utilize this fact to reduce the dimensionality of the unconditional multivariate Bernoulli distribution.

Because of the correlation of methylation rates between CpG sites, we treated methylation status $Y_{kis,j}$ at each CpG site j on short read sequence s as a branching process, taking advantage of the property of multivariate Bernoulli distribution [19]. We assumed that, for CpG site j , branching probabilities were the same for each short read sequence of all individuals in group k . Thus, we defined the branching probability $p_{kj} = P(Y_{kis,j} = 1 | Y_{kis,j-1} = 1)$ as the probability of methylated

sequence read at CpG site j , conditional on the methylated sequence read at CpG site $j - 1$ on the same short read sequence of the same individual. Similarly, we defined the branching probability $q_{kj} = P(Y_{kis,j} = 1 | Y_{kis,j-1} = 0)$ as the same probability, conditional on unmethylated sequence read at CpG site $j - 1$.

The methylation status $(Y_{kis,a}, Y_{kis,a+1}, \dots, Y_{kis,a+c-1})$ were generated as follows:

For the first CpG site of the sequence, the methylation status y_{kisa} was generated from Bernoulli distribution $Bern(m_a)$, with $m_a = (p_{ka} + q_{ka})/2$.

The methylation status $y_{kis,j}$ for $j = a + 1, \dots, a + c - 1$ was generated with $y_{kis,j} \sim Bern(p_{kj})$ if $y_{kis,j-1} = 1$ or $y_{kis,j} \sim Bern(q_{kj})$ if $y_{kis,j-1} = 0$.

After generating all the sequences at every CpG site for each individual, we calculated the total numbers of methylated and unmethylated short read sequences at CpG site j for individual i in group k , $\sum_s (y_{kis,j} = 1)$ and $\sum_s (y_{kis,j} = 0)$. Then the methylation count and the sequencing coverage are given by $m_{kij} = \sum_s (y_{kis,j} = 1)$ and $c_{kij} = \sum_s (y_{kis,j} = 1) + \sum_s (y_{kis,j} = 0)$, respectively.

We generated one CpG region with 24 CpG sites for each individual, 6 of which (from site 10 to 15) constituting the DMR. We simulated four groups of severity levels, with sample size 50 in each group, and repeated it with sample size 100. The branching probabilities, p_{kj} , were pre-determined. Also, we chose $q_{kj} = p_{kj} - 0.2$. We also simulated two different scenarios of DMR patterns.

Under Scenario 1, we chose the probabilities, p_{kj} , to be symmetric around the middle of the DMR (CpG sites 12 and 13). The predetermined probabilities p_{kj} and their symmetric pattern under Scenario 1 are presented in Table 1.

Table 1. Conditional probabilities p_{kj} at each CpG site for simulation of BFM under Scenario 1.

Site	1	2	...	9	10	11	12	13	14	15	16	17	...	24
group 1	0.44	0.46	...	0.6	0.62	0.64	0.66	0.66	0.64	0.62	0.6	0.58	...	0.44
group 2	0.44	0.46	...	0.6	0.72	0.74	0.76	0.76	0.74	0.72	0.6	0.58	...	0.44
group 3	0.44	0.46	...	0.6	0.82	0.84	0.86	0.86	0.84	0.82	0.6	0.58	...	0.44
group 4	0.44	0.46	...	0.6	0.92	0.94	0.96	0.96	0.94	0.92	0.6	0.58	...	0.44

Under Scenario 2, we randomly chose the CpG sites with the peak values of p_{kj} within the simulated DMR (between sites 10 and 15), varying it for different individuals. Specifically, for each individual in each group, we first generated a random number r (between 10 and 15) for the location of the CpG site with the highest methylation, and then chose the branching probabilities p_{kj} to increase from 10 to r and then decrease from r to 15. The p_{kj} for the non-DMCs remained the same as in Scenario 1. The second scenario is a more realistic depiction of the real world. However, the results and conclusions should be the same under both situations.

3. Results

3.1. Simulation Results

A total of 1000 replicates were simulated. For each replicate, the Bayes factor was calculated for each moving window with window size of 6. Calculations were based on 3000 Gibbs samplers, with 1000 Gibbs samplers for the burn-in period. The results of simulation for both the scenarios are presented in Table 2. As expected, the results are very similar for both scenarios. The results of Scenario 1 are plotted in Figures 1 and 2. As evident from Table 2, following Jeffreys' rule, when the moving windows contain at least three of the six CpG sites, we have strong evidence of differential methylation when sample size of 50 in each group and decisive evidence when sample size is 100.

Table 2. Mean Bayes factors at each CpG site, based on simulation studies.

Start	End	N = 50 (Scenario 1)	N = 100 (Scenario 1)	N = 50 (Scenario 2)
1	6	1.02	1.02	1.03
2	7	1.01	1.02	1.01

3	8	1.01	1.02	1.02
4	9	1.02	1.01	1.01
5	10	1.24	1.53	1.26
6	11	1.78	3.12	1.78
7	12	2.95	9.16	2.85
8	13	5.74	41.42	4.95
9	14	10.53	1052.07	9.31
10	15	18.79	8554.12	18.31
11	16	13.9	3718.77	13.79
12	17	8.44	306.07	8.12
13	18	4.43	21.91	4.5
14	19	2.4	5.66	2.6
15	20	1.52	2.22	1.6
16	21	1.07	1.11	1.07
17	22	1.03	1.04	1.02
18	23	1.01	1.03	1.02
19	24	1.03	1.03	1.01

All results show that the Bayes factors reach their maximum in the simulated DMR (CpG sites 10–15). However, the Bayes factors are not symmetric, the windows on the right side of the peak have larger values compared to those on the left side. This is attributed to the fact that the methylation status at a given site was generated conditional on that at the previous site of the same sequence. As expected, when the sample size is doubled the Bayes factors and the evidence in support of methylation increases significantly, as seen in Table 2 and Figures 1 and 2.

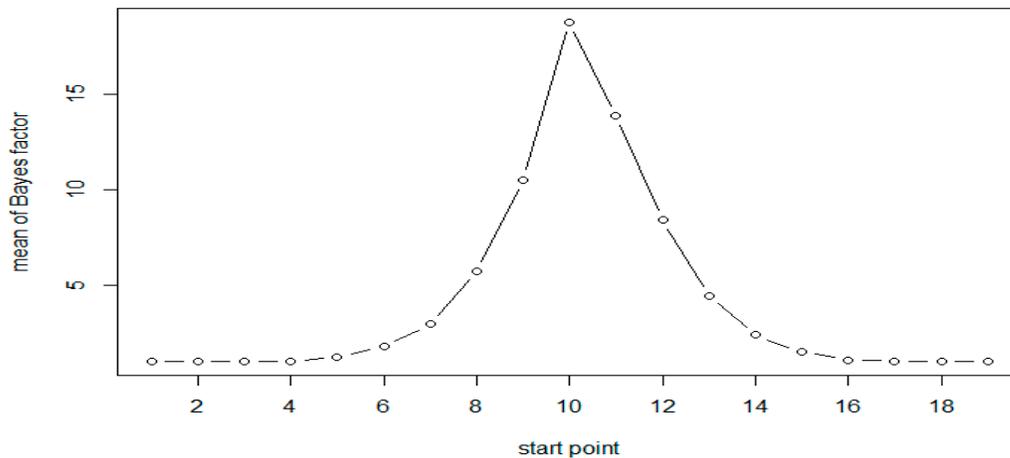


Figure 1. Mean of Bayes factors at each CpG site with $N = 50$ (Scenario 1).

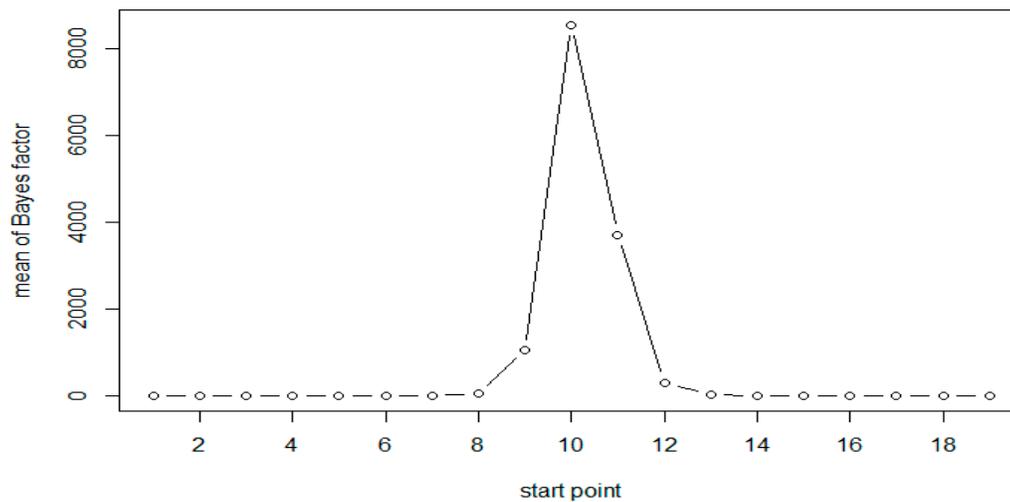


Figure 2. Mean of Bayes factors at each CpG site with $N = 100$ (Scenario 1).

In order to illustrate that our proposed method is statistically valid and to ensure that the BF in our method is a meaningful measure for comparison with frequentist approaches, we computed Bayes factors exclusively for all moving windows that do not include the differentially methylated sites 10–11. Among these Bayes factors, 95% were less than 1.34 and 99% were less than 1.50, both consistent with Jeffreys' rule. These values can be thought of as the cut-offs corresponding to 5% and 1% empirical type I error rates. We calculated the proportions of times the Bayes factors fall above these cut-offs, for all possible numbers of DMCs in the moving window. These results are given in Table 3. For the simulated data they are comparable to the conclusions based on frequentist interpretations of type I error and power. For the real data analysis, one could employ a permutation test to derive the cutoff values under the null hypothesis. However, since the frequentist interpretation is not necessarily consistent with the Bayesian conclusions, using Jeffrey's rule for decision making may be more desirable when analyzing real data.

Table 3. Proportions of Bayes factors that fell above the cut-off.

Cut-off point	Number of DMCs in the windows						
	0	1	2	3	4	5	6
1.34	0.050	0.56	0.97	1	1	1	1
1.5	0.010	0.35	0.91	1	1	1	1

3.2. Data analysis

We used our proposed BFM to analyze methylation data from a genome-wide association study of chronic lymphocytic leukemia (CLL), which manifests as a result of clonal expansion of malignant B cells. B-cell lymphoma, mostly prevalent among adults, is a heterogeneous disease [20,21]. It is clinically important to find heterogeneity of patients at the molecular level, which can help design specific interventions for patients at different severity levels.

Over the last decade, research in CLL has resulted in significant advances such as identification of several molecular alternations with prognostic values. These include specific cytogenetic patterns [22], mutational status of the immunoglobulin heavy chain variable gene (IgVH) [23] and expression of CD38 [24]. It has been found that patients lacking the mutation have a poorer prognosis. Patients with lower levels of CD38 have slower disease progression [23,25].

Several research groups have demonstrated that DNA methylation of multiple promoter-associated CpG islands is common in CLL [15,26,27]. Detection of aberrant DNA methylation in CLL

could result in the development of an epigenetic classification of the disease with prognostic and therapeutic potential.

CD19+ B cells from peripheral blood were collected from CLL samples and normal control subjects. All CLL samples were obtained from patients at the Ellis Fischel Cancer Center (EFCC), the Georgia Cancer Center of Augusta University and the North Shore-LIJ Health System in compliance with the local Institutional Review Boards [28].

Illumina sequencing reads were generated for each sample by using RRBS [29]. In total, 20–30 million reads were sequenced for each sample, and 63%–75% were successfully mapped to either strand of the human genome (hg18) [28]. The average sequencing depth per CpG was between 32x and 43x. Eventually RRBS provided counts of DNA molecules that were methylated or unmethylated at each CpG site, and overall methylation status of approximately 1.8–2.3 million CpG sites were determined consistently for each sample in the study [28].

Tong et al. [30] pointed out that aberrant DNA methylation associated with CLL were located more frequently on chromosome 19. Hence, we analyzed genome-wide methylation data on 17,917 CpG sites on Chromosome 19 of 40 patients.

3.3. Comparison of Bayesian Method with Scan Statistic Method for Two Groups

First, we tested for differential methylation under binary response, by dividing the samples into two groups based on CD38 level of 20 as the cut-off. We had 23 subjects with $CD38 \leq 20$ and 17 subjects with $CD38 > 20$. BFM and Scan statistic method (SSM) [31] were compared, using moving windows with 10 CpG sites in each window.

For comparing the two methods, we used a cut-off value of 2 for BFM and a 5% significance level for SSM. A total of 181 genes in DMRs were detected by SSM, and 183 genes were detected by BFM, using these criteria. Among these, 41 from SSM and 42 from BFM were found in PubMed publications as associated with leukemia (Table 4). There were 67 overlapping genes of which 18 were found in PubMed. They are ACP5, ATF5, BIRC8, C3, CARD8, CEACAM8, CERS1, CKM, CRTCL1, IL411, LAIR1, MAP1S, NFIX, PDE4C, PLEKHG2, PLVAP, RFX1, and ZNF331 [32–49].

Table 4. Comparison of BFM and SSM for window size of 10 ($p < 0.05$).

	BFM > 2	SSM ($p < 0.05$)	Common
Total	183	181	67
PubMed	42	41	18

C3 and LAIR1((INK4a))genes were both detected, which were shown to be related to acute myeloid leukemia [34,41]. Actually, both C3 and LAIR1 genes connect with the transcription factor CREB (cyclic AMP response element binding protein), which has a role in the pathogenesis of AML and other cancers [50,51].

3.4. Bayesian Method for Ordinal Group Responses

To test whether the methylation rates increase as the CD38 levels increase, the samples were classified into four risk groups based on CD38 level, with 5 non-leukemia subjects in group 1, 23 patients in group 2 with $CD38 \leq 20$, 9 patients in group 3 with $20 < CD38 \leq 50$, and 8 patients in group 4 with $CD38 > 50$. Though there are advantages of modeling CD38 as a continuous variable, but on the other hand, modeling as an ordinal variable is more robust to distributional assumptions. Again, moving windows of size of 10 were used for analysis. In fact, in clinical studies it is a common practice to put patients into discrete disease risk groups based on continuous measures.

Because of multiple testing issues associated with the comparison of four groups, we used a more stringent criterion of $BF > 19$ to evaluate the strength of evidence of differential methylation [8]. A total of 789 windows showed strong evidence of differential methylation using this criterion. The start and end positions in base pairs for each detected DMR were used in the UCSC genome browser to find the genes in the regions, and eventually 125 genes were found in these regions. Among them, 35 were associated with leukemia on PubMed literature. Some of these were not detected when only

two groups were considered even with a less stringent criterion. They are BRD4, ELL, ERCC1, ERCC2, GDF15, JUND, POLD1, PRDX2, RANBP3, SPIB and TSPAN16 [52–62].

4. Discussion

Results from our simulation study indicate that BFM is a valid approach to detect DMRs when considering ordinal group responses, since the calculated Bayes factors were very large for simulated DMRs, and close to 1 for non-DMRs. The real data analysis based on the CLL data also demonstrated that BFM is a valid method that is able to detect DMRs with methylation rates increasing (or decreasing) as disease severity increases.

In addition to being able to account for ordering of group responses, BFM also has an advantage of allowing for heterogeneity of methylation effects across CpG sites by modeling the methylation rates with a prior. Methods such as the SSM pools information across variants in a region, assuming that each CpG sites in the region have the same methylation rates.

BFM with mixed-effect regression, not only can allow for covariates, but also the correlation between CpG sites. It takes advantage of the flexibility of the Bayesian framework, including the use of prior information when available as well as computational convenience, and uses distributions such as the multivariate normal to incorporate the correlation structure with inverse Wishart distribution as the prior for the correlation matrix.

One disadvantage of the BFM is that it assumes that methylation rates of CpG sites within each moving window are independent of those outside of the window, while the SSM accounts for the correlation along the whole genome.

BFM used a moving window to help decide the location and length of DMRs. But practically, it is very difficult to know the exact length of DMRs. This limitation is very common in statistical genetics, not only for detecting DMRs, but also for detecting rare variants [63]. Cross validation or bootstrap approaches might help determine the window sizes. It could be possible to develop other methods, for example, using genes and promoters instead of moving windows, along with the BFM to detect DMRs.

As described by George and Laud [64], the Bayes factor used in the context of testing hypotheses is a meaningful measure of evidence because it is a reasonably approximate factor by which the odds are increased by the data. With default priors that are essentially flat over a wide range of the relevant parameter space, the approach is similar to the likelihood-based inference. However, direct comparison between methods such as BFM based on the Bayes factor with frequentist approaches should be done with caution, as the Bayes factor classification for decision process is not a precise calibration, but rather a descriptive statement about the standards of evidence. Our proposed method is rather exploratory in nature, leading to a ranked list of sites for follow up for formal confirmation.

We developed the BFM, focusing only on DNA methylation data. However, large-scale cancer genomics projects such as TCGA (The Cancer Genome Atlas Research Network) are currently generating multiple layers of genomics data for early tumor, including DNA copy number, methylation, and mRNA expression. Similar statistical methods for integrated analysis and systematic modeling of these genomics data deserve further attention.

Author Contributions: conceptualization, F.D., H.X. and V.G.; methodology, F.D., H.X, D.R., S.G. and V.G.; formal analysis, F.D.; writing—original draft preparation, F.D.; writing—H.X. and V.G.; data curation, H.S.

Funding: This research received no external funding.

Acknowledgment: The authors extend our special thanks and appreciation to one of the reviewers who provided a very extensive review of the manuscript with several insightful and constructive comments, which helped us improve the manuscript substantially.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Yokota, J. Tumor progression and metastasis. *Carcinogenesis* **2000**, *21*, 497–503.

2. Torre, L.A.; Siegel, R.L.; Jemal, A. Lung Cancer Statistics. In *Lung Cancer and Personalized Medicine: Current Knowledge and Therapies*; Ahmad, A., Gadgeel, S., Eds.; Advances in Experimental Medicine and Biology; Springer International Publishing: Cham, 2016; pp. 1–19 ISBN 978-3-319-24223-1.
3. Qureshi, S.A.; Bashir, M.U.; Yaqinuddin, A. Utility of DNA methylation markers for diagnosing cancer. *Int J Surg* **2010**, *8*, 194–198.
4. Varley, K.E.; Gertz, J.; Bowling, K.M.; Parker, S.L.; Reddy, T.E.; Pauli-Behn, F.; Cross, M.K.; Williams, B.A.; Stamatoyannopoulos, J.A.; Crawford, G.E.; et al. Dynamic DNA methylation across diverse human cell lines and tissues. *Genome Res.* **2013**, *23*, 555–567.
5. Klajic, J.; Fleischer, T.; Dejeux, E.; Edvardsen, H.; Warnberg, F.; Bukholm, I.; Lønning, P.E.; Solvang, H.; Børresen-Dale, A.-L.; Tost, J.; et al. Quantitative DNA methylation analyses reveal stage dependent DNA methylation and association to clinico-pathological factors in breast tumors. *BMC Cancer* **2013**, *13*, 456.
6. Watts, G.S.; Futscher, B.W.; Holtan, N.; Degeest, K.; Domann, F.E.; Rose, S.L. DNA methylation changes in ovarian cancer are cumulative with disease progression and identify tumor stage. *BMC Med Genomics* **2008**, *1*, 47.
7. Hoque, M.O. DNA methylation changes in prostate cancer: current developments and future clinical implementation. *Expert Rev. Mol. Diagn.* **2009**, *9*, 243–257.
8. Mitomi, H.; Fukui, N.; Tanaka, N.; Kanazawa, H.; Saito, T.; Matsuoka, T.; Yao, T. Aberrant p16 INK4a methylation is a frequent event in colorectal cancers: prognostic value and relation to mRNA expression and immunoreactivity. *J. Cancer Res. Clin. Oncol.* **2010**, *136*, 323–331.
9. Dunson, D.B.; Neelon, B. Bayesian Inference on Order-Constrained Parameters in Generalized Linear Models. *Biometrics* **2003**, *59*, 286–295.
10. Leek, J.T.; Scharpf, R.B.; Bravo, H.C.; Simcha, D.; Langmead, B.; Johnson, W.E.; Geman, D.; Baggerly, K.; Irizarry, R.A. Tackling the widespread and critical impact of batch effects in high-throughput data. *Nat. Rev. Genet.* **2010**, *11*, 733–739.
11. Bartholomew, D.J. A test of homogeneity for ordered alternatives. *Biometrika* **1959**, *46*, 36–48.
12. Robertson, T.; Wegman, E.J. Likelihood ratio tests for order restrictions in exponential families. *The Annals of Statistics* **1978**, 485–505.
13. Ayer, M.; Brunk, H.D.; Ewing, G.M.; Reid, W.T.; Silverman, E. An empirical distribution function for sampling with incomplete information. *The Annals of Mathematical Statistics* **1955**, 641–647.
14. Taylor, J.M.G.; Wang, L.; Li, Z. Analysis on binary responses with ordered covariates and missing data. *Stat Med* **2007**, *26*, 3443–3458.
15. Teschendorff, A.E.; Menon, U.; Gentry-Maharaj, A.; Ramus, S.J.; Weisenberger, D.J.; Shen, H.; Campan, M.; Nouchmeh, H.; Bell, C.G.; Maxwell, A.P.; et al. Age-dependent DNA methylation of genes that are suppressed in stem cells is a hallmark of cancer. *Genome Res.* **2010**, *20*, 440–446.
16. Kibriya, M.G.; Raza, M.; Jasmine, F.; Roy, S.; Paul-Brutus, R.; Rahaman, R.; Dodsworth, C.; Rakibuz-Zaman, M.; Kamal, M.; Ahsan, H. A genome-wide DNA methylation study in colorectal carcinoma. *BMC Med Genomics* **2011**, *4*, 50.
17. Jeffreys, H. *Theory of probability*, Clarendon; Oxford, 1961;
18. Kass, R.E.; Raftery, A.E. Bayes factors. *Journal of the American Statistical Association* **1995**, *90*, 773–795.
19. Dai, B.; Ding, S.; Wahba, G. Multivariate bernoulli distribution. *Bernoulli* **2013**, *19*, 1465–1483.
20. Chiorazzi, N.; Rai, K.R.; Ferrarini, M. Chronic lymphocytic leukemia. *N. Engl. J. Med.* **2005**, *352*, 804–815.
21. Keating, M.J.; Chiorazzi, N.; Messmer, B.; Damle, R.N.; Allen, S.L.; Rai, K.R.; Ferrarini, M.; Kipps, T.J. Biology and treatment of chronic lymphocytic leukemia. *Hematology Am Soc Hematol Educ Program* **2003**, 153–175.
22. Döhner, H.; Stilgenbauer, S.; Benner, A.; Leupolt, E.; Kröber, A.; Bullinger, L.; Döhner, K.; Bentz, M.; Lichter, P. Genomic aberrations and survival in chronic lymphocytic leukemia. *N. Engl. J. Med.* **2000**, *343*, 1910–1916.
23. Hamblin, T.J.; Davis, Z.; Gardiner, A.; Oscier, D.G.; Stevenson, F.K. Unmutated Ig V(H) genes are associated with a more aggressive form of chronic lymphocytic leukemia. *Blood* **1999**, *94*, 1848–1854.
24. Hamblin, T.J.; Orchard, J.A.; Gardiner, A.; Oscier, D.G.; Davis, Z.; Stevenson, F.K. Immunoglobulin V genes and CD38 expression in CLL. *Blood* **2000**, *95*, 2455–2457.
25. Damle, R.N.; Wasil, T.; Fais, F.; Ghiotto, F.; Valetto, A.; Allen, S.L.; Buchbinder, A.; Budman, D.; Dittmar, K.; Kolitz, J.; et al. Ig V gene mutation status and CD38 expression as novel prognostic indicators in chronic lymphocytic leukemia. *Blood* **1999**, *94*, 1840–1847.

26. Kanduri, M.; Cahill, N.; Göransson, H.; Enström, C.; Ryan, F.; Isaksson, A.; Rosenquist, R. Differential genome-wide array-based methylation profiles in prognostic subsets of chronic lymphocytic leukemia. *Blood* **2010**, *115*, 296–305.
27. Rahmatpanah, F.B.; Carstens, S.; Guo, J.; Sjahputera, O.; Taylor, K.H.; Duff, D.; Shi, H.; Davis, J.W.; Hooshmand, S.I.; Chitma-Matsiga, R.; et al. Differential DNA methylation patterns of small B-cell lymphoma subclasses with different clinical behavior. *Leukemia* **2006**, *20*, 1855–1862.
28. Pei, L.; Choi, J.-H.; Liu, J.; Lee, E.-J.; McCarthy, B.; Wilson, J.M.; Speir, E.; Awan, F.; Tae, H.; Arthur, G.; et al. Genome-wide DNA methylation analysis reveals novel epigenetic changes in chronic lymphocytic leukemia. *Epigenetics* **2012**, *7*, 567–578.
29. Meissner, A.; Gnirke, A.; Bell, G.W.; Ramsahoye, B.; Lander, E.S.; Jaenisch, R. Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis. *Nucleic Acids Res.* **2005**, *33*, 5868–5877.
30. Tong, W.-G.; Wierda, W.G.; Lin, E.; Kuang, S.-Q.; Bekele, B.N.; Estrov, Z.; Wei, Y.; Yang, H.; Keating, M.J.; Garcia-Manero, G. Genome-wide DNA methylation profiling of chronic lymphocytic leukemia allows identification of epigenetically repressed molecular pathways with clinical impact. *Epigenetics* **2010**, *5*, 499–508.
31. Dunbar, F.; Xu, H.; Ryu, D.; Ghosh, S.; Shi, H.; George, V. Computational Methods for Detection of Differentially Methylated Regions Using Kernel Distance and Scan Statistics. *Genes (Basel)* **2019**, *10*.
32. French, D.; Hamilton, L.H.; Mattano, L.A.; Sather, H.N.; Devidas, M.; Nachman, J.B.; Relling, M.V.; Children’s Oncology Group A PAI-1 (SERPINE1) polymorphism predicts osteonecrosis in children with acute lymphoblastic leukemia: a report from the Children’s Oncology Group. *Blood* **2008**, *111*, 4496–4499.
33. Wang, T.; Qian, D.; Hu, M.; Li, L.; Zhang, L.; Chen, H.; Yang, R.; Wang, B. Human cytomegalovirus inhibits apoptosis by regulating the activating transcription factor 5 signaling pathway in human malignant glioma cells. *Oncol Lett* **2014**, *8*, 1051–1057.
34. Glodkowska-Mrowka, E.; Solarzka, I.; Mrowka, P.; Bajorek, K.; Niesiobedzka-Krezel, J.; Seferynska, I.; Borg, K.; Stoklosa, T. Differential expression of BIRC family genes in chronic myeloid leukaemia—BIRC3 and BIRC8 as potential new candidates to identify disease progression. *Br. J. Haematol.* **2014**, *164*, 740–742.
35. Chae, H.-D.; Mitton, B.; Lacayo, N.J.; Sakamoto, K.M. Replication factor C3 is a CREB target gene that regulates cell cycle progression through the modulation of chromatin loading of PCNA. *Leukemia* **2015**, *29*, 1379–1389.
36. Xu, W.; Zhou, L.; Chen, Q.; Chen, C.; Fang, L.; Fang, X.; Shen, H. [Effect of YB-1 gene knockdown on human leukemia cell line K562/A02]. *Zhonghua Yi Xue Yi Chuan Xue Za Zhi* **2009**, *26*, 400–405.
37. Lasa, A.; Serrano, E.; Carricondo, M.; Carnicer, M.J.; Brunet, S.; Badell, I.; Sierra, J.; Aventín, A.; Nomdedéu, J.F. High expression of CEACAM6 and CEACAM8 mRNA in acute lymphoblastic leukemias. *Ann. Hematol.* **2008**, *87*, 205–211.
38. Camgoz, A.; Gencer, E.B.; Ural, A.U.; Baran, Y. Mechanisms responsible for nilotinib resistance in human chronic myeloid leukemia cells and reversal of resistance. *Leuk. Lymphoma* **2013**, *54*, 1279–1287.
39. Caldwell, M.K.; Digby, M.R.; Cameron-Smith, D. Short communication: Bovine-derived proteins activate STAT3 in human skeletal muscle in vitro. *J. Dairy Sci.* **2015**, *98*, 3016–3019.
40. Tang, H.-M.V.; Gao, W.-W.; Chan, C.-P.; Cheng, Y.; Deng, J.-J.; Yuen, K.-S.; Iha, H.; Jin, D.-Y. SIRT1 Suppresses Human T-Cell Leukemia Virus Type 1 Transcription. *J. Virol.* **2015**, *89*, 8623–8631.
41. Carbonnelle-Puscian, A.; Copie-Bergman, C.; Baia, M.; Martin-Garcia, N.; Allory, Y.; Haioun, C.; Crémades, A.; Abd-Alsamad, I.; Farcet, J.-P.; Gaulard, P.; et al. The novel immunosuppressive enzyme IL4I1 is expressed by neoplastic cells of several B-cell lymphomas and by tumor-associated macrophages. *Leukemia* **2009**, *23*, 952–960.
42. Kang, X.; Lu, Z.; Cui, C.; Deng, M.; Fan, Y.; Dong, B.; Han, X.; Xie, F.; Tyner, J.W.; Coligan, J.E.; et al. The ITIM-containing receptor LAIR1 is essential for acute myeloid leukaemia development. *Nat. Cell Biol.* **2015**, *17*, 665–677.
43. Haimovici, A.; Brigger, D.; Torbett, B.E.; Fey, M.F.; Tschan, M.P. Induction of the autophagy-associated gene MAP1S via PU.1 supports APL differentiation. *Leuk. Res.* **2014**, *38*, 1041–1047.
44. O’Connor, C.; Campos, J.; Osinski, J.M.; Gronostajski, R.M.; Michie, A.M.; Keeshan, K. Nfix expression critically modulates early B lymphopoiesis and myelopoiesis. *PLoS ONE* **2015**, *10*, e0120102.

45. Moon, E.; Lee, R.; Near, R.; Weintraub, L.; Wolda, S.; Lerner, A. Inhibition of PDE3B augments PDE4 inhibitor-induced apoptosis in a subset of patients with chronic lymphocytic leukemia. *Clin. Cancer Res.* **2002**, *8*, 589–595.
46. Runne, C.; Chen, S. PLEKHG2 promotes heterotrimeric G protein $\beta\gamma$ -stimulated lymphocyte migration via Rac and Cdc42 activation and actin polymerization. *Mol. Cell. Biol.* **2013**, *33*, 4294–4307.
47. Rantakari, P.; Auvinen, K.; Jäppinen, N.; Kapraali, M.; Valtonen, J.; Karikoski, M.; Gerke, H.; Iftakhar-E-Khuda, I.; Keuschnigg, J.; Umemoto, E.; et al. The endothelial protein PLVAP in lymphatics controls the entry of lymphocytes and antigens into lymph nodes. *Nat. Immunol.* **2015**, *16*, 386–396.
48. Chen, L.; Smith, L.; Johnson, M.R.; Wang, K.; Diasio, R.B.; Smith, J.B. Activation of protein kinase C induces nuclear translocation of RFX1 and down-regulates c-myc via an intron 1 X box in undifferentiated leukemia HL-60 cells. *J. Biol. Chem.* **2000**, *275*, 32227–32233.
49. McHale, C.M.; Zhang, L.; Lan, Q.; Li, G.; Hubbard, A.E.; Forrest, M.S.; Vermeulen, R.; Chen, J.; Shen, M.; Rappaport, S.M.; et al. Changes in the peripheral blood transcriptome associated with occupational benzene exposure identified by cross-comparison on two microarray platforms. *Genomics* **2009**, *93*, 343–349.
50. Crans-Vargas, H.N.; Landaw, E.M.; Bhatia, S.; Sandusky, G.; Moore, T.B.; Sakamoto, K.M. Expression of cyclic adenosine monophosphate response-element binding protein in acute leukemia. *Blood* **2002**, *99*, 2617–2619.
51. Mayr, B.; Montminy, M. Transcriptional regulation by the phosphorylation-dependent factor CREB. *Nat. Rev. Mol. Cell Biol.* **2001**, *2*, 599–609.
52. Stewart, H.J.S.; Horne, G.A.; Bastow, S.; Chevassut, T.J.T. BRD4 associates with p53 in DNMT3A-mutated leukemia cells and is implicated in apoptosis by the bromodomain inhibitor JQ1. *Cancer Med* **2013**, *2*, 826–835.
53. Muto, T.; Takeuchi, M.; Yamazaki, A.; Sugita, Y.; Tsukamoto, S.; Sakai, S.; Takeda, Y.; Mimura, N.; Ohwada, C.; Sakaida, E.; et al. Efficacy of myeloablative allogeneic hematopoietic stem cell transplantation in adult patients with MLL-ELL-positive acute myeloid leukemia. *Int. J. Hematol.* **2015**, *102*, 86–92.
54. Kong, J.H.; Mun, Y.-C.; Kim, S.; Choi, H.S.; Kim, Y.-K.; Kim, H.-J.; Moon, J.H.; Sohn, S.K.; Kim, S.-H.; Jung, C.W.; et al. Polymorphisms of ERCC1 genotype associated with response to imatinib therapy in chronic phase chronic myeloid leukemia. *Int. J. Hematol.* **2012**, *96*, 327–333.
55. Liu, D.; Wu, D.; Li, H.; Dong, M. The effect of XPD/ERCC2 Lys751Gln polymorphism on acute leukemia risk: a systematic review and meta-analysis. *Gene* **2014**, *538*, 209–216.
56. Secchiero, P.; Barbarotto, E.; Tiribelli, M.; Zerbini, C.; di Iasio, M.G.; Gonelli, A.; Cavazzini, F.; Campioni, D.; Fanin, R.; Cuneo, A.; et al. Functional integrity of the p53-mediated apoptotic pathway induced by the nongenotoxic agent nutlin-3 in B-cell chronic lymphocytic leukemia (B-CLL). *Blood* **2006**, *107*, 4122–4129.
57. Gazon, H.; Lemasson, I.; Polakowski, N.; Césaire, R.; Matsuoka, M.; Barbeau, B.; Mesnard, J.-M.; Peloponese, J.-M. Human T-cell leukemia virus type 1 (HTLV-1) bZIP factor requires cellular transcription factor JunD to upregulate HTLV-1 antisense transcription from the 3' long terminal repeat. *J. Virol.* **2012**, *86*, 9070–9078.
58. Sincennes, M.-C.; Humbert, M.; Grondin, B.; Lisi, V.; Veiga, D.F.T.; Haman, A.; Cazaux, C.; Mashtalir, N.; Affar, E.B.; Verreault, A.; et al. The LMO2 oncogene regulates DNA replication in hematopoietic cells. *Proc. Natl. Acad. Sci. U.S.A.* **2016**, *113*, 1393–1398.
59. Agrawal-Singh, S.; Isken, F.; Agelopoulos, K.; Klein, H.-U.; Thoennissen, N.H.; Koehler, G.; Hascher, A.; Bäumer, N.; Berdel, W.E.; Thiede, C.; et al. Genome-wide analysis of histone H3 acetylation patterns in AML identifies PRDX2 as an epigenetically silenced tumor suppressor gene. *Blood* **2012**, *119*, 2346–2357.
60. Hakata, Y.; Yamada, M.; Shida, H. A multifunctional domain in human CRM1 (exportin 1) mediates RanBP3 binding and multimerization of human T-cell leukemia virus type 1 Rex protein. *Mol. Cell. Biol.* **2003**, *23*, 8751–8761.
61. Talby, L.; Chambost, H.; Roubaud, M.-C.; N'Guyen, C.; Milili, M.; Lorigod, B.; Fossat, C.; Picard, C.; Gabert, J.; Chiappetta, P.; et al. The chemosensitivity to therapy of childhood early B acute lymphoblastic leukemia could be determined by the combined expression of CD34, SPI-B and BCR genes. *Leuk. Res.* **2006**, *30*, 665–676.
62. Juric, D.; Lacayo, N.J.; Ramsey, M.C.; Racevskis, J.; Wiernik, P.H.; Rowe, J.M.; Goldstone, A.H.; O'Dwyer, P.J.; Paietta, E.; Sikic, B.I. Differential gene expression patterns and interaction networks in BCR-ABL-positive and -negative adult acute lymphoblastic leukemias. *J. Clin. Oncol.* **2007**, *25*, 1341–1349.

63. Schaid, D.J.; Sinnwell, J.P.; McDonnell, S.K.; Thibodeau, S.N. Detecting genomic clustering of risk variants from sequence data: cases versus controls. *Hum. Genet.* **2013**, *132*, 1301–1309.
64. George, V.; Laud, P.W. A Bayesian approach to the transmission/disequilibrium test for binary traits. *Genetic epidemiology* **2002**, *22*, 41–51.



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).