cells

Article

Computational Detection of Breast Cancer Invasiveness with DNA Methylation Biomarkers

Chunyu Wang [1,*], Ning Zhao [2], Linlin Yuan [3] and Xiaoyan Liu [1,*]

[1] School of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001, China
[2] School of Life Science and Technology, Harbin Institute of Technology, Harbin 150080, China; zhaoning2016@hit.edu.cn
[3] College of Intelligence and Computing, Tianjin University, Tianjin 300350, China; yuan_linlin@163.com
* Correspondence: chunyu@hit.edu.cn (C.W.); liuxiaoyan@hit.edu.cn (X.L.); Tel.: +86-451-86413213 (C.W. & X.L.)

Received: 2 January 2020; Accepted: 30 January 2020; Published: 30 January 2020

Abstract: Breast cancer is the most common female malignancy. It has high mortality, primarily due to metastasis and recurrence. Patients with invasive and noninvasive breast cancer require different treatments, so there is an urgent need for predictive tools to guide clinical decision making and avoid overtreatment of noninvasive breast cancer and undertreatment of invasive cases. Here, we divided the sample set based on the genome-wide methylation distance to make full use of metastatic cancer data. Specifically, we implemented two differential methylation analysis methods to identify specific CpG sites. After effective dimensionality reduction, we constructed a methylation-based classifier using the Random Forest algorithm to categorize the primary breast cancer. We took advantage of breast cancer (BRCA) HM450 DNA methylation data and accompanying clinical data from The Cancer Genome Atlas (TCGA) database to validate the performance of the classifier. Overall, this study demonstrates DNA methylation as a potential biomarker to predict breast tumor invasiveness and as a possible parameter that could be included in the studies aiming to predict breast cancer aggressiveness. However, more comparative studies are needed to assess its usability in the clinic. Towards this, we developed a website based on these algorithms to facilitate its use in studies and predictions of breast cancer invasiveness.

Keywords: breast cancer; metastasis; invasiveness; DNA methylation

1. Introduction

According to a National Cancer Center report describing the status and trends of cancer in China in 2017, breast cancer is the most common female malignancy in the country. It is a complex and heterogeneous disease with multiple molecular subtypes, which can be defined by immunohistochemistry or microarray profiling [1–3]. The incidence of breast cancer is increasing, but there are limited curative options when metastasis develops. Mammography has been shown to be an economical non-invasive tool for early diagnosis [4]. However, the high mortality rate of breast cancer is, to a large extent, a result of metastasis, which affects up to 40% of women suffering from this disease [5]. Unfortunately, despite the continuous improvements of medical technology, we still cannot control cancer metastasis[6].

Metastasis refers to the process in which malignant tumor cells relocate and then continue to grow in other parts of the body separate from the primary site, by traveling through lymphatic channels, blood vessels, or the body cavity; it is often the main reason for the failure of tumor treatment [7–12]. The occurrence and progress of metastases in tumors are nonrandom and thus potentially predictable. For example, colorectal cancer typically spreads to the liver, while breast cancer primarily metastasizes to bone marrow and lung [13]. This emphasizes the need for novel prognostic tools to guide clinical decision-making about diagnosis and treatment. Predicting the invasiveness and progression of tumors is crucial for clinical decision-making and avoiding both overtreatment of indolent breast cancer and undertreatment of aggressive disease [14].

In view of this, several groups have analyzed the molecular expression profiles of primary tumors and metastases to regional lymph nodes or distant sites [15–19]. In recent years, large-scale sequencing techniques, such as next-generation sequencing and microarray [20–23], have enabled the systematic detection of abnormalities of the genome, transcriptome, and epigenome associated with cancer [22,24–34]. One can reevaluate the cancer progress through an integrative way [35] to understand their regulation system [36,37]. Moreover, this has been enhanced by single-cell level omics [38]. However, metastatic cancer-related data are still extremely scarce, which is a result of the complexity of metastasis. The recurrence and metastasis of tumors usually occur several years after the primary tumor has been removed [39]. To investigate the possibility of metastasis of a primary tumor occurring long after treatment, follow-up should be performed for a long time, which is expensive and laborious [40]. Moreover, as secondary cancer growing at new sites cannot be routinely removed like the primary tumor during treatment for metastatic breast cancer, the tissue is not available for research until an autopsy is performed [41].

Cancer-specific alterations of DNA methylation are closely related to a variety of malignancies [42–45]. Recent studies investigating the genomic lesions of primary and metastatic cancers revealed that some specific DNA methylation changes could account for tumor metastasis and progression [46,47]. The appearance of lymph node metastasis is an indication of tumor cells developing the ability to leave the primary site and spread to a new site; it thus acts as a marker of the ability of the tumor to establish distant metastases [48].

In this study, we used DNA methylation data for lymph node metastasis to study the invasiveness of breast cancer. To overcome the problems of a small amount of data and lack of samples for metastatic cases, we used a novel method to identify sample labels based on their DNA methylation markers and then constructed a classifier for identifying invasive breast cancer. Upon applying this classifier to The Cancer Genome Atlas (TCGA) BRCA samples, the acquired results were satisfactory.

2. Materials and Methods

Figure 1 provides an overview of the experimental procedure. It consists of four major components: (i) differential methylation analysis selecting specific CpG sites, (ii) filtering of redundant feature sites by dimensionality reduction, (iii) building classifiers based on the Random Forest algorithm, and (iv) evaluation of classifier performance by enrichment analysis.

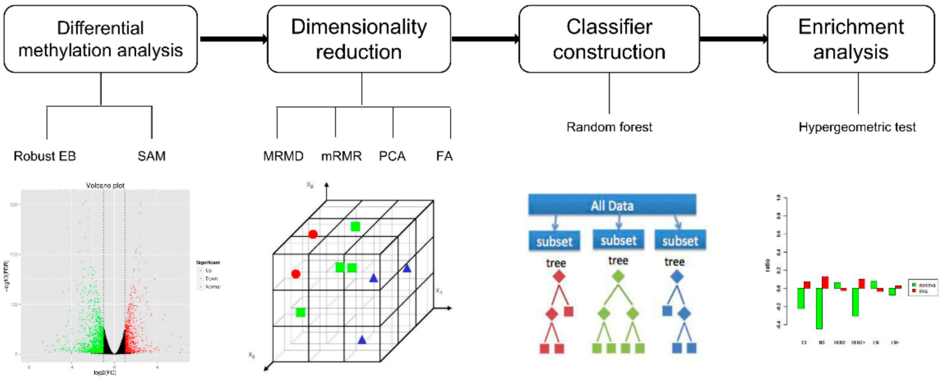

Figure 1. Schematic diagram of the presented stepwise analysis. Step 1, differential methylation analysis: Two methods were carried out to select the differential methylation sites. Step 2, dimensionality reduction: using four methods to reduce the dimension. Step 3, four kinds of dimensionality reduction results were used to construct the classifier respectively. Step 4, enrichment analysis: Hypergeometric test evaluated and compared the performances of the classifiers.

2.1. Materials: DNA Methylation Datasets

In this study, we used three genome-wide DNA methylation profiles of breast cancer. The first two datasets were acquired from the Gene Expression Omnibus (GEO). One of them includes 44 matched primary breast cancer and regional metastases (GSE58999) [49], while the other contains 80 primary breast cancer and 40 normal samples (GSE66695) [50]. Another dataset was obtained from The Cancer Genome Atlas (TCGA); this dataset comprises data on a primary breast cancer population ($n = 766$) and a normal population ($n = 97$). The DNA methylation profile was measured by Infinium HumanMethylation450 BeadChip (HM450), which contains more than 480,000 probes. The HM450 DNA methylation profile was also used, which covers 99% of the NCBI reference genes and can provide data on DNA methylation at single-base resolution [51].

In the pretreatment processing, we deleted single-nucleotide polymorphism (SNP) probes and CpH (A/T/C) sites and then removed CpG sites at which more than 30% of samples had missing values (NA). The remaining missing values were complemented by the *k*-nearest neighbor, using the "impute.knnl" function in the R package "limma."

2.2. Methods: Study Design

We calculated Euclidean distances between any two matched samples using all 393,806 probes, and then removed abnormal sample pairs with excessive distances between them using the method of quartile. Subsequently, we obtained 40 matched primary breast cancer and lymph node metastatic sample pairs whose clinical information was obtained from GSE58999 and in which primary breast cancer samples were defined as invasive. Furthermore, in order to determine whether the primary breast cancer samples in GSE66695 are invasive, we calculated Euclidean distances between lymph node metastatic samples and these primary breast cancer samples using all probes. We believe that the greater the distance between the sample and the lymph node metastatic samples the more likely it is to be noninvasive. Therefore, if the minimum distance between a sample and the 40 lymph node metastases was 10 units larger than the maximum distance of the paired sample, we defined the sample as noninvasive. In this way, we identified 20 noninvasive samples. Invasive (40 samples) and noninvasive (20 samples) samples, together with 40 normal samples, collectively constituted the training sets.

2.2.1. Robust Empirical Bayes

Empirical Bayes (EB) is a statistical algorithm that assumes a Bayesian hierarchical model for the variances and estimates the prior distribution from the marginal distribution of the observed data [52]. Robust EB improves differential methylation analysis by strengthening the hyperparameter estimation procedure and achieves robustness with regard to inaccurate working priors by conditioning on the rank of each estimated log-fold change rather than on the actual observation. It uses log-expression values to fit linear models for each CpG site; computes its moderated *t*-statistic, moderated *F*-statistic, and log-odds of differential methylation; and selects differential methylation sites by implementing robust hyperparameter estimation [53]. This method was applied using the R package "limma" [54].

2.2.2. Significance Analysis of Microarrays

Significance Analysis of Microarrays (SAM) establishes a *d*-statistic for each site, correlating these sites with an outcome variable, such as metastasis, after which CpG sites are rearranged in descending order of *d*-statistic. SAM simulates a null distribution by permutating the mark of the group randomly to calculate the *p*-value of the difference in methylation between groups [55]. The permutation algorithm rearranges and repartitions samples. This process is performed M times, and for site i, its statistic of pth permutation is recorded as $d_p(i)$. The sites are rearranged in descending order of $d_p(i)$. Note that:

$$d_E(i) = \sum_p \frac{d_p(i)}{M} \tag{1}$$

$$\Delta = \left| d(i) - d_E(i) \right| \tag{2}$$

Here, screening is performed for differential methylation sites as those with Δ greater than the threshold. This method was carried out using the R package "samr" [56].

2.2.3. Maximum Relevance Maximum Distance

Maximum Relevance Maximum Distance (MRMD) is a Java-based feature selection method [57]. It aims to select features with maximum relevance and maximum distance. It uses Pearson's correlation coefficient to measure the correlation of feature and label, and uses Euclidean distance between features to calculate redundancy, which was also widely used in clustering [58–60]. MRMD ranks all candidate features based on the calculated Pearson's correlation coefficient and Euclidean distance, then constructs a simple classifier using the top-ranked features, and finally selects a feature list with the best classification accuracy. MRMD can pick out the optimal number of features, having the lowest redundancy and the strongest correlation with the categorical variable.

2.2.4. Minimal Redundancy Maximal Relevance

Minimum Redundancy Maximum Relevance (mRMR) is a filtered feature selection method [61]. mRMR is based on the concept of minimizing the correlation between different features, while maximizing the correlation between features and target classes. Each correlation is measured based on mutual information. Such mutual information can be regarded as the amount of information about another random variable contained in a random variable, which is a measure of the statistical correlation between two random variables [62]. After using mRMR for feature selection, the ranking of each feature regarding its importance is obtained. Next, cross-validation is performed to select the subset of features with the best performance [63–66].

2.2.5. Principal Component Analysis

Principal Component Analysis (PCA) is one of the most widely used data compression algorithms. It maps n-dimensional data to k-dimensional space by linear projection ($k < n$) and obtains the new data with the largest variance in the projected dimension; it results in fewer data dimensions being used and more features of the original data being retained [67]. The principal components are selected based on the cumulative percentage of total variation. We chose to retain the number of principal components representing more than 95% of the total variation, which is a frequently used threshold [68]. We implemented PCA using the Dimensionality Reduction feature of scikit-learn [69].

2.2.6. Factor Analysis

Factor Analysis (FA) is a statistical technique of extracting common factors from a variable population [70]. The common factor refers to the inherent hidden factor between different variables. FA is based on the concept of classifying the observed variables; highly correlated variables are grouped into the same class, while variables in different categories are poorly correlated. Each type of variable actually represents a basic structure, a common factor. In this way, most of the information of the original data can be reflected by a few factors, which enables the data to be condensed. FA was carried out using the Dimensionality Reduction part of scikit-learn [69].

2.2.7. Unsupervised Hierarchical Clustering

Heatmaps were used to display the difference of DNA methylation levels between invasive and noninvasive groups. We randomly chose 10 samples from each of the two patient groups and applied the "levelplot" function implemented in the R package "lattice" to visualize their difference [71]. The unsupervised hierarchical clustering was performed with the hclust function in R (method = "complete").

2.2.8. Constructing the DNA Methylation-Based Invasiveness Classifier

After differential methylation analysis and dimensionality reduction, we generated a list of 134 variably methylated CpG sites between the invasive and noninvasive groups and a list of 14 variably methylated CpG sites between normal and cancer groups. We used these sites as features to construct classifiers. The Random Forest model is a successful ensemble learning classifier. It can build a series of classification and decision trees through training and has been proven to perform well for the classification of multi-gene microarray chip data [55]. We therefore built classifiers based on a random forest algorithm and used 10-fold cross-validation to evaluate the accuracy of the model.

2.2.9. Hypergeometric Test

Hypergeometric tests were used to determine whether sample groups were particularly associated with some clinical factors. The use of a hypergeometric distribution is a common method of enrichment analysis [72]. It calculated the probability of enrichment of a clinical factor for each class of samples, taking the enrichment ratio of a whole sample set in this clinical factor as the background. The *p*-values of multiple-test correction were adjusted using the FDR method.

3. Results

3.1. Feature Selection

Differential methylation analyses were performed using the methods of Robust EB and SAM. For the results of these two methods, we set thresholds of logFC ≥ 1.5 (fold change) and *p*-value < 0.01 to select differentially methylated sites. Robust EB identified 8653 and 11,808 CpG sites in the cancer–normal group and invasiveness–noninvasiveness group, respectively. SAM selected 14,096 and 7329 CpG sites in the cancer–normal group and invasiveness–noninvasiveness group, respectively. The false discovery rate (FDR) of both methods was set to 0.01. To further reduce the false positive rate, we selected the results that overlapped between Robust EB and SAM. Using the overlapping results, there were 7888 and 6461 sites that were differentially methylated between the two groups (the detailed information on CpGs is available in the Supplementary Materials: Table S1).

In the process of tumor occurrence and development, methylation alert occurs in a genome-wide scale, and many changes are consistent. Therefore, there are many redundant features. In this context, further dimensionality reduction is required to filter out redundant CpG sites. Here, we used four dimensionality reduction methods to select the optimal number of features, namely, MRMD, mRMR, PCA, and FA. MRMD and mRMR are feature selection methods, which identify CpG sites that are related to tags but not related to each other. PCA and FA combine the original features into new features to achieve dimensionality reduction and lose as little of the information conveyed in the original data as possible. To compare the effects of the different dimensionality reduction methods, we constructed classifiers with the four sets of associated results. The dimensionality reduction results and classifier accuracy of the four methods are listed in Table 1.

Table 1. The number of selected CpG sites and the performance of four classifiers for DNA methylation profiles.

		Normal	Invasiveness
MRMD	Number of CpG	14	134
	Training Accuracy	97%	93.6%
	Testing Accuracy	96.9%	549/217
mRMR	Number of CpG	12	5
	Training Accuracy	99%	100%
	Testing Accuracy	96.9%	611/165

Table 1. *Cont.*

		Normal	**Invasiveness**
PCA	Number of CpG	8	3
	Training Accuracy	99%	95%
	Testing Accuracy	91%	454/312
FA	Number of CpG	80	60
	Training Accuracy	99%	93.3%
	Testing Accuracy	94.8%	664/102

3.2. Development of Classifier Based on DNA Methylation Biomarkers

Next, we established a DNA methylation-based BRCA invasiveness classifier to categorize primary breast cancer as either invasive or noninvasive using the four dimensionality reduction results. We used the Random Forest algorithm to train the classifier and confirmed the validity of the classification by 10-fold cross-validation. With regard to training accuracy, the four results were all satisfactory. The accuracy of the normal prediction was as high as 99%, and the prediction accuracy of invasion was as high as 95%. The classification results are listed in detail in Table 1. To provide more useful information, we also added a false positive (FP) and false negative (FN) in the Supplementary Materials (Table S2). However, the performance of the classifier is more reflected in the test accuracy. For this reason, we downloaded an independent dataset from TCGA database to test the classifiers.

3.3. TCGA Beast Cancer Cohort Confirms the Performance of Classifiers

To test the classifier on a new dataset, we used the publicly available breast cancer (BRCA) HM450 DNA methylation data and accompanying clinical data from The Cancer Genome Atlas (TCGA) project. We tested 766 primary breast cancer samples and 97 normal samples using the classifiers. The accuracy of the four classifiers for normal samples was as high as 96.9%, almost as high as that for the training. However, for the prediction of invasiveness, the four dimensionality reduction methods provided markedly different results.

To evaluate the predictive performance of the classifier, we investigated some metastasis-related clinical indicators of samples from the two groups of prediction. We applied the hypergeometric test to confirm the significance of the enrichment of prediction samples in some clinical indicators, such as primary tumor stage (T-stage), regional lymph node stage (N-stage), human epidermal growth factor receptor 2 (HER2) status, and presence of lymph node metastasis (LN+) in pathological report. T3 indicates that the maximum diameter of the primary breast tumor is more than 5 cm, while N3 refers to lymph node metastases of ipsilateral internal mammary and ipsilateral. HER2 positivity (HER2+) suggests that breast cancer is susceptible to relapse or metastases. LN+ indicates that lymph node metastases are mentioned in the pathological report. These indicators all reflect greater invasiveness [73]. According to the prediction labels of the four classifiers, we applied the hypergeometric test on these indicators separately. Figure 2 shows comparisons of the enrichment ratio and significance of the four results. The histogram above the X axis represents the enrichment ratio of the sample population on the clinical feature, which is greater than 1. We expect to see more enrichment of the sample predicted to be invasive on clinical factors suggesting metastasis. In terms of the dimensionality reduction results of MRMD, samples predicted to be invasive were significantly enriched for indicators suggesting susceptibility to tumor metastasis, such as T3, N3, and HER2+. The detection of LN+ is associated with the number of examined lymph nodes, and the number of examined lymph nodes is related to the scope of the lymph node dissection during the operation. There are some other pathways to metastasize, such as direct infiltration and blood transfer [74]. These factors probably lead to inaccuracy of measuring invasiveness with LN+ and account for the poor significance of IN+. Two indicators were extremely significant under the effect of PCA, and two were significant in the result of FA. These results all demonstrate that our classifier is effective.

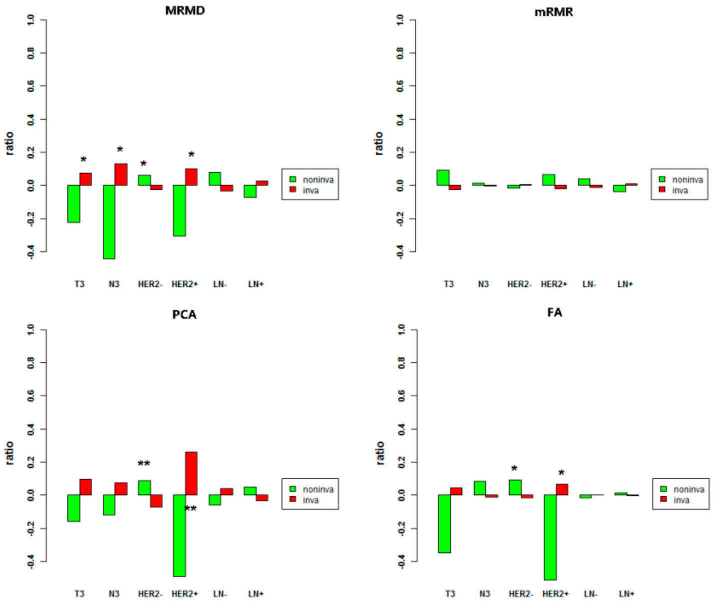

Figure 2. Four enrichment analyses results of clinical indicators in the two tumor clusters. Red indicates samples predicted by the classifier as invasive, and green indicates samples predicted to be non-invasive. The X axis represents clinical indicators; the Y axis represents the enrichment ratio of Table 0. ** is extremely significant ($p < 0.01$).

3.4. Methylation Differences between the Invasive and Noninvasive Groups

Alterations in DNA methylation occur in all cancers, play important roles in the development and progression of cancer, and are associated with tumor aggressiveness in most types of cancer [42]. In this study, we identified noninvasive primary tumor samples by the degree of dissimilarity of DNA methylation between primary cancer and lymph node metastases. Next, we applied two statistical methods to search for probes that were differentially methylated between the invasive and noninvasive groups. Based on the overlapping results between robust EB and SAM, 6461 CpG sites were revealed. An excessive number of features would increase the complexity of the model and lead to overfitting. We therefore further reduced the dimensions to filter out redundant CpG sites and used the optimal number of features to build classifiers. Finally, we evaluated the performance of these classifiers through enrichment analysis. In terms of the biological analysis results, MRMD had the best dimensionality reduction effect, which retained 134 differentially methylated sites (Supplementary Materials: Table S3). Therefore, we further explored these 134 CpG sites.

To investigate whether this method of sample division and feature selection can effectively identify CpG sites associated with metastasis, we performed unsupervised hierarchical clustering on 20 randomly selected samples based on their methylation levels on these 134 sites. The heatmap shows the results of unsupervised clustering and the difference of methylation patterns between the two types of sample. As shown in Figure 3, the two samples were completely separated and the methylation levels also showed significant differences.

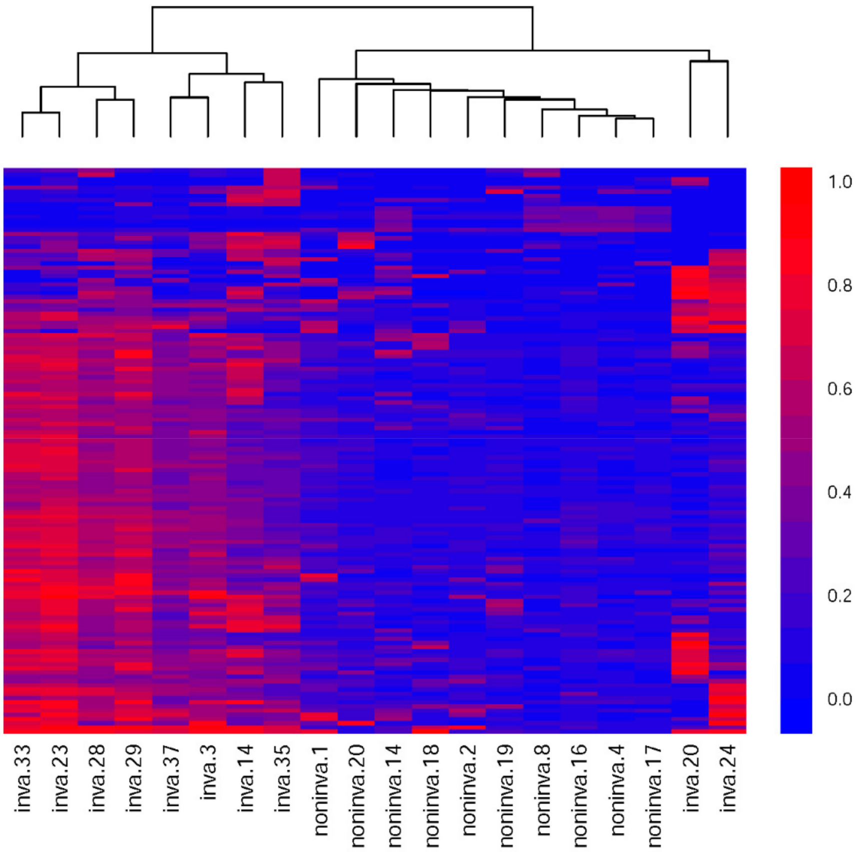

Figure 3. Unsupervised clustering and heatmap of 20 samples based on the 134 differently methylated probes. Each column corresponds to a sample and each row corresponds to a CpG site. Color indicates methylation value. With color ranging from blue to red, methylation values range from small to large. Color key is to the right.

3.5. Genes Related to Metastasis

We further listed the genes that the 134 CpG sites correspond to, giving a total of 98 (Supplementary Materials: Table S4). The HCMDB database annotated most of these genes as being associated with metastasis [75]. Researchers systematically reviewed more than 7000 published papers in the PubMed database using "metastasis" and corresponding gene symbols as the keywords, and manually curated 2183 genes related to metastasis. Comparing the 98 genes with these metastatic cancer-related genes confirmed in the literature, we found that 12 genes were associated with metastatic cancer, which were confirmed in the literature, among which five were associated with breast cancer metastasis. Table 2 lists the 12 confirmed metastasis-related genes and associated literature annotations, with black indicating the genes related to breast cancer metastasis. This indicates that our method can effectively detect CpG methylation sites related to cancer metastasis and also provides new information for the discovery of more relevant genes.

Taking into account the dimensional reduction results of MRMR, only five sites are needed to completely separate the training set. We also analyzed the corresponding genes for these five sites. The results show that two of the sites are on the gene body, while the remaining three are currently not annotated. We suspect that these five sites could become biomarkers suggestive of breast cancer metastasis. In view of this, we have listed information on these five sites in the Supplementary Materials (Table S5).

Table 2. Known metastasis-associated genes and their descriptions in literatures.

Gene	Description
ABCC5	*ABCC5* functions as a mediator of breast cancer skeletal metastasis.
ASCL2	Functions as a suppressor of colorectal cancer metastasis by down-regulating the *ASCL2-CXCR4* signaling axis.
BNIP3	"*BNIP3* deletion can be used as a prognostic marker of tumor progression to metastasis in human triple-negative breast cancer"
FLI1	"This study for the first time identifies *FLI1* as a clinically and functionally important target gene of metastasis, providing a rationale for developing *FLI1* inhibitors in the treatment of breast cancer."
ITGA6	"The role of *PTHrP* in breast cancer growth and metastasis may be mediated via upregulation of integrin alpha6beta4 expression and Akt activation, with consequent inactivation of *GSK*-3."
MPL	"In migrating cancer stem cells isolated from primary human colorectal cancers, *CD110*(+) and *CDCP1*(+) subpopulations mediate organ-specific lung and liver metastasis."
NCOR2	"Thyroid hormone receptors induce TRAIL expression, and TRAIL thus synthesized acts in concert with simultaneously synthesized Bcl-xL to promote metastasis"
RHOB	"*RHOB* belongs to a novel class of ""genes of recurrence"" that have a dual role in metastasis and treatment resistance."
SLITRK3	*SLITRK3* expression is a highly significant predictor of gastrointestinal stromal tumor recurrence and metastasis.
SND1	"*SND1* is a novel *MTDH*-interacting protein and has shown that it is a functionally and clinically significant mediator of metastasis."
TRPS1	"*TRPS1* plays a crucial role in osteosarcoma angiogenesis, metastasis and clinical surgical stage."
WWOX	*WWOX* is associated with tumorigenicity and metastasis of head and neck and gastric signet-ring cell carcinoma.

Gene is gene symbol; " … " This is a direct quote from the corresponding literature.

3.6. Website BMMP

To facilitate research on breast cancer metastasis, we developed a website for the prediction of invasiveness of breast cancer, BMMP (BRCA methylation metastasis prediction [76]). BMMP is a Java-based website that uses the classification model of MRMD's dimensionality reduction results. BMMP enables users to predict the invasiveness of breast cancer by pasting or uploading DNA methylation profiling data. It also lists experimentally validated metastasis-associated genes used by the classifier. The data for each step of the experiment and classification model are accessible on the website.

4. Discussion and Conclusions

Invasive and noninvasive breast cancers differ markedly in their clinical manifestations and prognosis. They also have different clinical treatments. Accurately identifying the differences and predicting tumor invasiveness can have a radical impact on breast cancer research. As an important epigenetic regulator, DNA methylation can serve as a stable marker for samples. In the study reported here, we inferred whether tumors are invasive or noninvasive based on the DNA methylation pattern of breast cancer.

We used two differential methylation analysis methods to identify the CpG sites differentially methylated between the two groups. After reducing the dimensions of these features, we constructed a methylation-based invasiveness classifier to categorize primary breast cancer as either invasive or noninvasive. Finally, we confirmed the credibility of the classifier by comparing the extent to which some clinical factors were particularly enriched in the predicted samples. Our study provides molecular-based support for determining the invasiveness of breast cancer and indicates the potential impact of applying this approach to clinical decision-making. Although this method was only used to evaluate the invasiveness of breast cancer in this study, we believe it should be generally applicable to other types of tumor.

Although this method can guide the study of breast cancer metastasis, it has some limitations. For example, breast cancer is a highly heterogeneous disease, so a fixed classifier may have different prediction accuracies for samples from different subclasses. Considering this, we chose samples including four disease subtypes as a training set [49]. As the study of heterogeneity deepens, research on metastatic cancer has begun to consider this issue, leading to the development of a personalized committee classifier [77], for example. We expect that more intense research can further contribute to revealing the molecular mechanisms involved in breast cancer metastasis and potentially help in diagnosing and treating it. Furthermore, link prediction [78–82], probabilistic models [83–86] and computational intelligence methods [20,87–90], which have been successfully applied in many areas [63,91–100], can be considered in BRCA methylation metastasis prediction.

Supplementary Materials: The following are available online at http://www.mdpi.com/2073-4409/9/2/326/s1, Table S1: The detailed information on all CpGs; Table S2: The detailed prediction performance; Table S3: 134 differentially methylated sites detected from MRMD; Table S4: 98 genes related to the 134 CpG markers; Table S5: Five breast cancer biomarkers CpGs predicted by our method.

Author Contributions: Conceptualization, C.W.; Data curation, N.Z. and L.Y.; Methodology, C.W. and X.L.; Writing–original draft, C.W.; Writing–review & editing, N.Z. and L.Y. All authors have read and agreed to the published version of the manuscript

Funding: This research was funded by the National Natural Science Foundation of China (No. 61872114, No. 61532014, No. 61671189), and the National Key Research and Development Plan of China (No. 2016YFC0901902).

Acknowledgments: We thank Liwen Bianji, Edanz Group China (www.liwenbianji.cn/ac), for editing the English text of a draft of this manuscript.

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

References

1. Perou, C.M.; Sørlie, T.; Eisen, M.B.; Van, d.R.M.; Jeffrey, S.S.; Rees, C.A.; Pollack, J.R.; Ross, D.T.; Johnsen, H.; Akslen, L.A. Molecular portraits of human breast tumors. *Nature* **2012**, *490*, 747–752. [CrossRef]
2. Sorlie, T.; Tibshirani, R.; Parker, J.; Hastie, T.; Marron, J.S.; Nobel, A.; Deng, S.; Johnsen, H.; Pesich, R.; Geisler, S. Repeated observation of breast tumor subtypes in independent gene expression data sets. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 8418–8423. [CrossRef] [PubMed]
3. Nones, K.; Johnson, J.; Newell, F.; Patch, A.; Thorne, H.; Kazakoff, S.; de Luca, X.; Parsons, M.; Ferguson, K.; Reid, L. Whole-genome sequencing reveals clinically relevant insights into the aetiology of familial breast cancers. *Ann. Oncol.* **2019**, *30*, 1071–1079. [CrossRef]
4. Wang, J.H.; Yang, X.; Cai, H.M.; Tan, W.C.; Jin, C.Z.; Li, L. Discrimination of Breast Cancer with Microcalcifications on Mammography by Deep Learning. *Sci. Rep.* **2016**, *6*, 1–9. [CrossRef]
5. Doornebal, C.W.; Klarenbeek, S.; Braumuller, T.M.; Klijn, C.N.; Ciampricotti, M.; Hau, C.S.; Hollmann, M.W.; Jonkers, J.; Visser, K.E.D. A Preclinical Mouse Model of Invasive Lobular Breast Cancer Metastasis. *Cancer Res.* **2013**, *73*, 353–363. [CrossRef] [PubMed]
6. Kozłowski, J.; Kozłowska, A.; Kocki, J. Breast cancer metastasis - insight into selected molecular mechanisms of the phenomenon. *Postępy Hig. I Med. Doświadczalnej* **2014**, *69*, 447–451. [CrossRef]
7. Fingleton, B. Molecular targets in metastasis: Lessons from genomic approaches. *Cancer Genom. Proteom.* **2007**, *4*, 211–221.
8. Fokas, E.; Engenhart-Cabillic, R.; Daniilidis, K.; Rose, F.; An, H.X. Metastasis: The seed and soil theory gains identity. *Cancer Metastasis Rev.* **2007**, *26*, 705–715. [CrossRef]
9. Hanahan, D.; Weinberg, R.A. The hallmark of cancer. *Cell* **2000**, *100*, 57–71. [CrossRef]
10. Poste, G.; Fidler, I.J. The pathogenesis of cancer metastasis. *Nature* **1980**, *283*, 139–146. [CrossRef]
11. Du, X.Q.; Li, X.R.; Li, W.; Yan, Y.T.; Zhang, Y.P. Identification and Analysis of Cancer Diagnosis Using Probabilistic Classification Vector Machines with Feature Selection. *Curr. Bioinform.* **2018**, *13*, 625–632. [CrossRef]
12. Liu, H.; Luo, L.B.; Cheng, Z.Z.; Sun, J.J.; Guan, J.H.; Zheng, J.; Zhou, S.G. Group-sparse Modeling Drug-kinase Networks for Predicting Combinatorial Drug Sensitivity in Cancer Cells. *Curr. Bioinform.* **2018**, *13*, 437–443. [CrossRef]

13. Ring, B.Z.; Ross, D.T. Predicting the sites of metastases. *Genome Biol.* **2005**, *6*, 241. [CrossRef]
14. Ma, Q.; Reeves, J.H.; Liberles, D.A.; Yu, L.; Chang, Z.; Zhao, J.; Cui, J.; Xu, Y.; Liu, L. A phylogenetic model for understanding the effect of gene duplication on cancer progression. *Nucleic Acids Res.* **2013**, *42*, 2870–2878. [CrossRef]
15. Ellsworth, R.E.; Seebach, J.; Field, L.A.; Heckman, C.; Kane, J.; Hooke, J.A.; Love, B.; Shriver, C.D. A gene expression signature that defines breast cancer metastases. *Clin. Exp. Metastasis* **2009**, *26*, 205–213. [CrossRef]
16. Feng, Y.; Sun, B.; Li, X.; Zhang, L.; Niu, Y.; Xiao, C.; Ning, L.; Fang, Z.; Wang, Y.; Zhang, L. Differentially expressed genes between primary cancer and paired lymph node metastases predict clinical outcome of node-positive breast cancer patients. *Breast Cancer Res. Treat.* **2007**, *103*, 319–329. [CrossRef]
17. Hao, X.; Sun, B.; Hu, L.; Lähdesmäki, H.; Dunmire, V.; Feng, Y.; Zhang, S.W.; Wang, H.; Wu, C.; Wang, H. Differential gene and protein expression in primary breast malignancies and their lymph node metastases as revealed by combined cDNA microarray and tissue microarray analysis. *Cancer* **2010**, *100*, 1110–1122. [CrossRef]
18. Suzuki, M.; Tarin, D. Gene expression profiling of human lymph node metastases and matched primary breast carcinomas: Clinical implications. *Mol. Oncol.* **2008**, *1*, 172–180. [CrossRef]
19. Weigelt, B.; Glas, A.M.; Wessels, L.F.; Witteveen, A.T.; Peterse, J.L.; van't Veer, L.J. Gene expression profiles of primary breast tumors maintained in distant metastases. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 15901–15905. [CrossRef]
20. Ren, G.H.; Cao, Y.T.; Wen, S.P.; Huang, T.W.; Zeng, Z.G. A modified Elman neural network with a new learning rate scheme. *Neurocomputing* **2018**, *286*, 11–18. [CrossRef]
21. Cheng, L.; Jiang, Y.; Ju, H.; Sun, J.; Peng, J.; Zhou, M.; Hu, Y. InfAcrOnt: Calculating cross-ontology term similarities using information flow by a random walk. *Bmc Genom.* **2018**, *19*, 919. [CrossRef]
22. Cheng, L.; Hu, Y.; Sun, J.; Zhou, M.; Jiang, Q. DincRNA: A comprehensive web-based bioinformatics toolkit for exploring disease associations and ncRNA function. *Bioinformatics* **2018**, *34*, 1953–1956. [CrossRef] [PubMed]
23. Lin, M.; Li, X.; Guo, H.; Ji, F.; Ye, L.; Ma, X.; Cheng, W. Identification of Bone Metastasis-associated Genes of Gastric Cancer by Genome-wide Transcriptional Profiling. *Curr. Bioinform.* **2019**, *14*, 62–69. [CrossRef]
24. Bianchini, G.; Iwamoto, T.; Qi, Y.; Coutant, C.; Shiang, C.Y.; Wang, B.; Santarpia, L.; Valero, V.; Hortobagyi, G.N.; Symmans, W.F. Prognostic and therapeutic implications of distinct kinase expression patterns in different subtypes of breast cancer. *Cancer Res.* **2010**, *70*, 8852. [CrossRef] [PubMed]
25. Xin, Z.; Ma, Q.; Ren, S.; Wang, G.; Li, F. The understanding of circular RNAs as special triggers in carcinogenesis. *Brief. Funct. Genom.* **2017**, *16*, 80–86. [CrossRef] [PubMed]
26. Breiman, L. Bagging Predictors. *Mach. Learn.* **1996**, *24*, 123–140. [CrossRef]
27. Xu, Y.; Wang, Y.; Luo, J.; Zhao, W.; Zhou, X. Deep learning of the splicing (epi)genetic code reveals a novel candidate mechanism linking histone modifications to ESC fate decision. *Nucleic Acids Res.* **2017**, *45*, 12100–12112. [CrossRef]
28. Xu, M.Z.; Zhao, Z.M.; Zhang, X.P.; Gao, A.Q.; Wu, S.Y.; Wang, J.Y. Synstable Fusion: A Network-Based Algorithm for Estimating Driver Genes in Fusion Structures. *Molecules* **2018**, *23*, 2055. [CrossRef]
29. Cheng, L.; Wang, P.; Tian, R.; Wang, S.; Guo, Q.; Luo, M.; Zhou, W.; Liu, G.; Jiang, H.; Jiang, Q. LncRNA2Target v2.0: A comprehensive database for target genes of lncRNAs in human and mouse. *Nucleic Acids Res.* **2019**, *47*, D140–D144. [CrossRef]
30. Tang, W.; Wan, S.; Yang, Z.; Teschendorff, A.E.; Zou, Q. Tumor origin detection with tissue-specific miRNA and DNA methylation markers. *Bioinformatics* **2018**, *34*, 398–406. [CrossRef]
31. Liao, Z.J.; Li, D.P.; Wang, X.R.; Li, L.S.; Zou, Q. Cancer Diagnosis Through IsomiR Expression with Machine Learning Method. *Curr. Bioinform.* **2018**, *13*, 57–63. [CrossRef]
32. Zeng, W.; Wang, F.; Ma, Y.; Liang, X.C.; Chen, P. Dysfunctional Mechanism of Liver Cancer Mediated by Transcription Factor and Non-coding RNA. *Curr. Bioinform.* **2019**, *14*, 100–107. [CrossRef]
33. Liu, G.; Jiang, Q. Alzheimer's disease CD33 rs3865444 variant does not contribute to cognitive performance. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, E1589–E1590. [CrossRef]
34. Liu, G.; Zhao, Y.; Jin, S.; Hu, Y.; Wang, T.; Tian, R.; Han, Z.; Xu, D.; Jiang, Q. Circulating vitamin E levels and Alzheimer's disease: A Mendelian randomization study. *Neurobiol. Aging* **2018**, *72*, 189.e181–189.e189. [CrossRef]

35. Xu, A.D.; Chen, J.Z.; Peng, H.; Han, G.Q.; Cai, H.M. Simultaneous Interrogation of Cancer Omics to Identify Subtypes With Significant Clinical Differences. *Front. Genet.* **2019**, *10*, 17. [CrossRef] [PubMed]
36. Chen, J.; Han, G.; Xu, A.; Cai, H. Identification of Multidimensional Regulatory Modules through Multi-graph Matching with Network Constraints. *IEEE Trans. Bio-Med. Eng.* **2019**. [CrossRef]
37. Jiang, Q.; Wang, J.; Wang, Y.; Ma, R.; Wu, X.; Li, Y. TF2LncRNA: Identifying common transcription factors for a list of lncRNA genes from ChIP-Seq data. *Biomed. Res. Int* **2014**, *2014*, 317642. [CrossRef] [PubMed]
38. Xu, Y.; Zhou, X. Applications of Single-Cell Sequencing for Multiomics. *Methods Mol. Biol* **2018**, *1754*, 327–374. [CrossRef] [PubMed]
39. Kalimutho, M.; Nones, K.; Srihari, S.; Duijf, P.H.; Waddell, N.; Khanna, K.K. Patterns of genomic instability in breast cancer. *Trends Pharmacol. Sci.* **2019**, *40*, 198–211. [CrossRef] [PubMed]
40. Duijf, P.H.; Nanayakkara, D.; Nones, K.; Srihari, S.; Kalimutho, M.; Khanna, K.K. Mechanisms of genomic instability in breast cancer. *Trends Mol. Med.* **2019**, *25*, 595–611. [CrossRef]
41. Mundbjerg, K.; Chopra, S.; Alemozaffar, M.; Duymich, C.; Lakshminarasimhan, R.; Nichols, P.W.; Aron, M.; Siegmund, K.D.; Ukimura, O.; Aron, M. Identifying aggressive prostate cancer foci using a DNA methylation classifier. *Genome Biol.* **2017**, *18*, 3. [CrossRef] [PubMed]
42. Hatada, I. The Epigenomics of Cancer. In *An Omics Perspective on Cancer Research*; Cho, W., Ed.; Springer: Dordrecht, The Netherlands, 2010; pp. 51–67. [CrossRef]
43. Cui, J.; Yin, Y.; Ma, Q.; Wang, G.; Olman, V.; Zhang, Y.; Chou, W.C.; Hong, C.S.; Zhang, C.; Cao, S. Comprehensive characterization of the genomic alterations in human gastric cancer. *Int. J. Cancer* **2015**, *137*, 86–95. [CrossRef] [PubMed]
44. Nones, K.; Waddell, N.; Song, S.; Patch, A.M.; Miller, D.; Johns, A.; Wu, J.; Kassahn, K.S.; Wood, D.; Bailey, P. Genome-wide DNA methylation patterns in pancreatic ductal adenocarcinoma reveal epigenetic deregulation of SLIT-ROBO, ITGA2 and MET signaling. *Int. J. Cancer* **2014**, *135*, 1110–1118. [CrossRef] [PubMed]
45. Wang, G.; Luo, X.; Wang, J.; Wan, J.; Xia, S.; Zhu, H.; Qian, J.; Wang, Y. MeDReaders: A database for transcription factors that bind to methylated DNA. *Nucleic Acids Res.* **2018**, *46*, D146–D151. [CrossRef] [PubMed]
46. Chiam, K.; Ricciardelli, C.; Bianco-Miotto, T. Epigenetic biomarkers in prostate cancer: Current and future uses. *Cancer Lett.* **2014**, *342*, 248–256. [CrossRef]
47. Vitale, A.M.; Matigian, N.A.; Cristino, A.S.; Nones, K.; Ravishankar, S.; Bellette, B.; Fan, Y.; Wood, S.A.; Wolvetang, E.; Mackay-Sim, A. DNA methylation in schizophrenia in different patient-derived cell types. *npj Schizophrenia* **2017**, *3*, 1–11. [CrossRef]
48. Fisher, B.; Bauer, M.; Wickerham, D.L.; Redmond, C.K.; Fisher, E.R.; Cruz, A.B.; Foster, R.; Gardner, B.; Lerner, H.; Margolese, R. Relation of number of positive axillary nodes to the prognosis of patients with primary breast cancer. An NSABP update. *Cancer* **2015**, *52*, 1551–1557. [CrossRef]
49. Reyngold, M.; Turcan, S.; Giri, D.; Kannan, K.; Walsh, L.A.; Viale, A.; Drobnjak, M.; Vahdat, L.T.; Lee, W.; Chan, T.A. Remodeling of the Methylation Landscape in Breast Cancer Metastasis. *PLoS ONE* **2014**, *9*, e103896. [CrossRef]
50. Jones, L.R.; Young, W.; Divine, G.; Datta, I.; Chen, K.M.; Ozog, D.; Worsham, M.J. Genome-Wide Scan for Methylation Profiles in Breast Cancer. *Dis. Markers* **2015**, *2015*, 943176. [CrossRef]
51. Sandoval, J.; Heyn, H.; Moran, S.; Serra-Musach, J.; Pujana, M.A.; Bibikova, M.; Esteller, M. Validation of a DNA methylation microarray for 450,000 CpG sites in the human genome. *Epigenetics* **2016**, *6*, 692–702. [CrossRef]
52. Phipson, B.; Lee, S.; Majewski, I.J.; Alexander, W.S.; Smyth, G.K. Robust hyperparameter estimation protects against hypervariable genes and improves power to detect differential expression. *Ann. Appl. Stat.* **2016**, *10*, 946. [CrossRef] [PubMed]
53. Smyth, G.K. Linear Models and Empirical Bayes Methods for Assessing Differential Expression in Microarray Experiments. *Stat. Appl Genet. Mol. Biol.* **2004**, *3*, Article3. [CrossRef] [PubMed]
54. Ritchie, M.E.; Phipson, B.; Wu, D.; Hu, Y.; Law, C.W.; Shi, W.; Smyth, G.K. Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **2015**, *43*, e47. [CrossRef] [PubMed]
55. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]

56. Jolma, A.; Yan, J.; Whitington, T.; Toivonen, J.; Nitta, K.R.; Rastas, P.; Morgunova, E.; Enge, M.; Taipale, M.; Wei, G.; et al. DNA-binding specificities of human transcription factors. *Cell* **2013**, *152*, 327–339. [CrossRef] [PubMed]
57. Zou, Q.; Zeng, J.; Cao, L.; Ji, R. A novel features ranking metric with application to scalable visual and bioinformatics data classification. *Neurocomputing* **2016**, *173*, 346–354. [CrossRef]
58. Xu, Y.; Guo, M.; Liu, X.; Wang, C.; Liu, Y.; Liu, G. Identify bilayer modules via pseudo-3D clustering: Applications to miRNA-gene bilayer networks. *Nucleic Acids Res.* **2016**, *44*, e152. [CrossRef]
59. Cheng, L.; Sun, J.; Xu, W.Y.; Dong, L.X.; Hu, Y.; Zhou, M. OAHG: An integrated resource for annotating human genes with multi-level ontologies. *Sci. Rep.* **2016**, *6*, 1–9. [CrossRef]
60. Cheng, L.; Jiang, Y.; Wang, Z.; Shi, H.; Sun, J.; Yang, H.; Zhang, S.; Hu, Y.; Zhou, M. DisSim: An online system for exploring significant similar diseases and exhibiting potential therapeutic drugs. *Sci. Rep.* **2016**, *6*, 30024. [CrossRef]
61. Peng, H.; Long, F.; Ding, C. Feature Selection Based on Mutual Information: Criteria of Max-Dependency, Max-Relevance, and Min-Redundancy. *Ieee Trans. Pattern Anal. Mach. Intell.* **2005**, *27*, 1226–1238. [CrossRef]
62. Ding, C.; Peng, H. Minimum Redundancy Feature Selection from Microarray Gene Expression Data. *J. Bioinform. Comput. Biol.* **2005**, *3*, 185–205. [CrossRef]
63. Tang, Y.; Liu, D.; Wang, Z.; Wen, T.; Deng, L. A boosting approach for prediction of protein-RNA binding residues. *Bmc Bioinform.* **2017**, *18*, 465. [CrossRef]
64. Liu, B. BioSeq-Analysis: A platform for DNA, RNA and protein sequence analysis based on machine learning approaches. *Brief. Bioinform.* **2017**, *20*, 1280–1294. [CrossRef]
65. Dao, F.Y.; Lv, H.; Wang, F.; Feng, C.Q.; Ding, H.; Chen, W.; Lin, H. Identify origin of replication in Saccharomyces cerevisiae using two-step feature selection technique. *Bioinformatics* **2019**, *35*, 2075–2083. [CrossRef]
66. Chen, W.; Feng, P.; Liu, T.; Jin, D. Recent advances in machine learning methods for predicting heat shock proteins. *Curr. Drug Metab.* **2018**, *20*, 224–228. [CrossRef]
67. Tipping, M.E.; Bishop, C.M. Probabilistic Principal Component Analysis. *J. R. Stat. Soc.* **1999**, *61*, 611–622. [CrossRef]
68. Minka, T.P. Automatic choice of dimensionality for PCA. In Proceedings of the International Conference on Neural Information Processing Systems, Montreal, QC, Canada, 29 December 2000; pp. 577–583.
69. Pedregosa, F.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; Vanderplas, J. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2013**, *12*, 2825–2830.
70. Gorsuch, R.L. *Factor Analysis*; Lawrence Erlbaum Associates: Mahwah, NJ, USA, 1983.
71. Chen, S.; Doolen, G.D. Lattice Boltzmann method for fluid flows. *Annu. Rev. Fluid Mech.* **1998**, *30*, 329–364. [CrossRef]
72. Johnson, N.L.; Kotz, S.; Kemp, A.W. *Univariate Discrete Distributions*; John Wiley & Sons: Hoboken, NJ, USA, 1992.
73. Haffner, M.C.; Mosbruger, T.; Esopi, D.M.; Fedor, H.; Heaphy, C.M.; Walker, D.A.; Adejola, N.; Gürel, M.; Hicks, J.; Meeker, A.K. Tracking the clonal origin of lethal prostate cancer. *J. Clin. Investig.* **2013**, *123*, 4918–4922. [CrossRef]
74. Marino, N.; Woditschka, S.; Reed, L.T.; Nakayama, J.; Mayer, M.; Wetzel, M.; Steeg, P.S. Breast Cancer Metastasis. *Am. J. Pathol.* **2013**, *183*, 1084–1095. [CrossRef]
75. Zheng, G.; Ma, Y.; Zou, Y.; Yin, A.; Li, W.; Dong, D. HCMDB: The human cancer metastasis database. *Nucleic Acids Res.* **2018**, *46*, D950–D955. [CrossRef]
76. Wang, C.; Yuan, L. BRCA Methylation Metastasis Prediction. Available online: http://server.malab.cn/BMMP/ (accessed on 30 January 2020).
77. Jahid, M.J.; Huang, T.H.; Ruan, J. A personalized committee classification approach to improving prediction of breast cancer metastasis. *Bioinformatics* **2014**, *30*, 1858–1866. [CrossRef]
78. Zeng, X.X.; Liu, L.; Lu, L.Y.; Zou, Q. Prediction of potential disease-associated microRNAs using structural perturbation method. *Bioinformatics* **2018**, *34*, 2425–2432. [CrossRef]
79. Ding, Y.; Tang, J.; Guo, F. Identification of drug-target interactions via multiple information integration. *Inf. Sci.* **2017**, *418*, 546–560. [CrossRef]
80. Xiao, Y.; Zhang, J.; Deng, L. Prediction of lncRNA-protein interactions using HeteSim scores based on heterogeneous networks. *Sci. Rep.* **2017**, *7*, 3664. [CrossRef]

81. Zhang, X.; Zou, Q.; Rodriguez-Paton, A.; Zeng, X.X. Meta-Path Methods for Prioritizing Candidate Disease miRNAs. *IEEE-Acm Trans. Comput. Biol. Bioinform.* **2019**, *16*, 283–291. [CrossRef] [PubMed]
82. Jiang, Q.; Wang, G.; Jin, S.; Li, Y.; Wang, Y. Predicting human microRNA-disease associations based on support vector machine. *Int J. Data Min. Bioinform* **2013**, *8*, 282–293. [CrossRef] [PubMed]
83. Guo, F.; Li, S.C.; Du, P.; Wang, L. Probabilistic Models for Capturing More Physicochemical Properties on Protein-Protein Interface. *J. Chem. Inf. Modeling* **2014**, *54*, 1798–1809. [CrossRef]
84. Ding, Y.; Tang, J.; Guo, F. Identification of Protein-Ligand Binding Sites by Sequence Information and Ensemble Classifier. *J. Chem. Inf. Modeling* **2017**, *57*, 3149–3161. [CrossRef] [PubMed]
85. Wang, G.; Wang, Y.; Feng, W.; Wang, X.; Yang, J.Y.; Zhao, Y.; Wang, Y.; Liu, Y. Transcription factor and microRNA regulation in androgen-dependent and -independent prostate cancer cells. *Bmc Genom.* **2008**, *9* Suppl 2, S22. [CrossRef]
86. Wang, G.; Wang, Y.; Teng, M.; Zhang, D.; Li, L.; Liu, Y. Signal transducers and activators of transcription-1 (STAT1) regulates microRNA transcription in interferon gamma-stimulated HeLa cells. *PLoS ONE* **2010**, *5*, e11794. [CrossRef]
87. Cabarle, F.G.C.; Adorna, H.N.; Jiang, M.; Zeng, X. Spiking Neural P Systems With Scheduled Synapses. *IEEE Trans. Nanobioscience* **2017**, *16*, 792–801. [CrossRef] [PubMed]
88. Dong, M.H.; Wen, S.P.; Zeng, Z.G.; Yan, Z.; Huang, T.W. Sparse fully convolutional network for face labeling. *Neurocomputing* **2019**, *331*, 465–472. [CrossRef]
89. Li, Z.L.; Dong, M.H.; Wen, S.P.; Hu, X.; Zhou, P.; Zeng, Z.G. CLU-CNNs: Object detection for medical images. *Neurocomputing* **2019**, *350*, 53–59. [CrossRef]
90. Zhao, Y.; Wang, F.; Juan, L. MicroRNA Promoter Identification in Arabidopsis Using Multiple Histone Markers. *Biomed. Res. Int* **2015**, *2015*, 861402. [CrossRef] [PubMed]
91. Zou, Q.; Li, J.; Song, L.; Zeng, X.; Wang, G. Similarity computation strategies in the microRNA-disease network: A Survey. *Brief. Funct. Genom.* **2016**, *15*, 55–64. [CrossRef]
92. Zeng, X.; Ding, N.; Rodríguezpatón, A.; Quan, Z.J.B.M.G. Probability-based collaborative filtering model for predicting gene–disease associations. *Bmc Med. Genom.* **2017**, *10*, 76. [CrossRef]
93. Ding, Y.; Tang, J.; Guo, F. Identification of Protein-Protein Interactions via a Novel Matrix-Based Sequence Representation Model with Amino Acid Contact Information. *Int. J. Mol. Sci.* **2016**, *17*, 1623. [CrossRef]
94. Cheng, L.; Yang, H.; Zhao, H.; Pei, X.; Shi, H.; Sun, J.; Zhang, Y.; Wang, Z.; Zhou, M. MetSigDis: A manually curated resource for the metabolic signatures of diseases. *Brief. Bioinform.* **2017**, *20*, 203–209. [CrossRef]
95. Pan, Y.W.; Wang, Z.; Zhan, W.; Deng, L. Computational identification of binding energy hot spots in protein-RNA complexes using an ensemble approach. *Bioinformatics* **2018**, *34*, 1473–1480. [CrossRef]
96. Liu, G.; Jin, S.; Hu, Y.; Jiang, Q. Disease status affects the association between rs4813620 and the expression of Alzheimer's disease susceptibility gene TRIB3. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, E10519–E10520. [CrossRef]
97. Liu, G.; Hu, Y.; Han, Z.; Jin, S.; Jiang, Q. Genetic variant rs17185536 regulates SIM1 gene expression in human brain hypothalamus. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 3347–3348. [CrossRef]
98. Wang, X.; Yu, B.; Ma, A.; Chen, C.; Liu, B.; Ma, Q. Protein–protein interaction sites prediction by ensemble random forests with synthetic minority oversampling technique. *Bioinformatics* **2018**, *35*, 2395–2402. [CrossRef]
99. Liu, G.; Wang, T.; Tian, R.; Hu, Y.; Han, Z.; Wang, P.; Zhou, W.; Ren, P.; Zong, J.; Jin, S.; et al. Alzheimer's Disease Risk Variant rs2373115 Regulates GAB2 and NARS2 Expression in Human Brain Tissues. *J. Mol. Neurosci.* **2018**, *66*, 37–43. [CrossRef]
100. Liu, G.; Xu, Y.; Jiang, Y.; Zhang, L.; Feng, R.; Jiang, Q. PICALM rs3851179 Variant Confers Susceptibility to Alzheimer's Disease in Chinese Population. *Mol. Neurobiol.* **2017**, *54*, 3131–3136. [CrossRef]