Hindawi
Contrast Media & Molecular Imaging
Volume 2021, Article ID 8202068, 8 pages
https://doi.org/10.1155/2021/8202068

Research Article

The Vista of Application of Specific Anaphylaxis Accurate Diagnosis Based on DNA Single-Nucleotide Methylation Sites

Xiangjie Guo,[1,2] **Yaqin Bai**,[1] **Hualin Guo**,[1] **Peng Wu**,[1] **Hao Li**,[1] **Liqin Zhai**,[3,4] **Yan Feng**,[4] **Jianguo Li**,[2] **Cairong Gao**,[1] **and Keming Yun**[1]

[1]*Department of Forensic Medicine, Shanxi Medical University, Taiyuan, ShanXi, China*
[2]*Shanxi Key Laboratory of Drug Toxicology and Drug for Radiation Injury, China Institute for Radiation Protection, Taiyuan, ShanXi, China*
[3]*Pathologic Department of People's Hospital of Shanxi Province, Taiyuan, Shanxi, China*
[4]*Department of Otolaryngology Head & Neck Surgery, The First Hospital Shanxi Medical University, Taiyuan, Shanxi, China*

Correspondence should be addressed to Xiangjie Guo; 258187101@qq.com, Jianguo Li; 996436323@qq.com, Cairong Gao; gaocairong5175@163.com, and Keming Yun; best001@sxmu.edu.cn

Received 22 September 2021; Revised 21 October 2021; Accepted 29 October 2021; Published 24 November 2021

Academic Editor: Yuvaraja Teekaraman

Anaphylaxis has rapidly spread around the world in the last several decades. Environmental factors seem to play a major role, and epigenetic marks, especially DNA methylation, get more attention. We discussed several GEO opening data classifications with TOP 100 specific methylation region values (normalized M-values on line) by machine learning, which are remarkable to classify specific anaphylaxis after monoallergen exposure. Then, we sequenced the whole-genome DNA methylation of six people (3 wormwood monoallergen atopic rhinitis patients and 3 normal-immune people) during the pollen season and analyzed the difference of the single nucleotide and DNA region. The results' divergences were obvious (the differential single nucleotides were mostly distributed in nongene regions but the differential DNA regions of GWAS, on the other hand), which may have caused most single nucleotides to be concealed in the regions' sequences. Therefore, we suggest that we should conduct more "pragmatic" and directly find special single-nucleotide changes after exposure to atopic allergens instead of complex correlativity. It is possible to try to use DNA methylation marks to accurately diagnose anaphylaxis and form a machine learning classification based on the single methylated CpGs.

1. Introduction

In the past few decades, the allergic disease incidence rate has increased yearly at a rapid rate, and these changes are much faster than the genome. The incidence of anaphylaxis ranges from 1 to 761 per 100000 person-years for total anaphylaxis and 1 to 77 per 100 000 person-years for food-induced anaphylaxis worldwide [1]. In Taiwan, the incidence rate of anaphylaxis has increased at an average rate of 5% annually since 2001 [2]. Allergic rhinitis is the most common chronic allergic disease, and its incidence is rising in parallel with other IgE-mediated diseases, affecting 10 to 30% of adults and up to 40% of children [3, 4]. In some Western developed countries, food-induced anaphylaxis already seems to be an epidemic (and highest in children) [5, 6]. However, among the Asian population, the incidence of drug-induced anaphylaxis increased faster compared with other types. Also, in some developing countries, such as Brazil, anaphylactic shock has high incidence rate [7]. Different age groups focused on distinct allergic incentives, and the incidence of allergies increased over time [8]. The large-scale intervention trials for food allergy support that the decrease in early exposure to allergens will increase the risk of food allergy [6]. The distinguishable type of allergy incidence varies significantly in different countries, even within countries, suggesting environmental factors play a major role compared with genetic factors in these changes [9]. The rapid development of the global economy, upturn

living standards, and lifestyle changes are accelerating the increased number of allergic cases and new allergens, which is a challenge for accurate diagnosis in forensic medicine and the medical domain.

1.1. The Relation of Epigenetics and Allergy. The rapid global outbreak of anaphylaxis is inevitable. To date, genome-wide linkage and association studies have identified many allergy-associated genes or loci [10–12]. However, the pattern of genetic susceptibility cannot explain all the risks of anaphylactic raise. There is evidence that the risk of allergy is higher in mothers than in fathers, and it is hereditable [13]. Because epigenetic marks are also heritable and capture responses to environmental factors [14, 15], it is logical that epigenetics plays an important role in the event of allergy. Furthermore, epigenome changes can be altered by many environmental exposures and often lead to rapid and persistent changes in gene expression [16]. Though monozygotic twins with the same genetic background are discordant for allergic rhinitis, they differ in peripheral blood mononuclear cell (PBMC) gene expression levels, and the sensitization of familiar allergens differs because of environmental contributions [17, 18]. In some studies, it has been determined that the influences can modify allergic patients' gene expression through DNA methylation [14, 19–21]. Thus, DNA methylation as an epigenetic mark represents a logical way to reflect allergy disease conditions.

1.2. The Association between DNA Methylation and Gene. DNA methylation refers to the covalent bonding of a methyl group to the 5th carbon position of the cytosine of the genomic CpG dinucleotide under the action of DNA methyltransferase. DNA methylation is observed in different sequences; however, it is almost exclusively found in CpG dinucleotides in humans. There are CpG-rich sequences termed CpG islands (CGI), which are generally unmethylated [22] and associated with histone modifications such as H3K4me [23, 24], but dissociative CpGs are methylated in general. There are around 50,000 CGIs in the human genome and more dissociative CpGs [25]. High CpG methylation in genomes increases the frequency of spontaneous mutations because methylated C residues spontaneously deaminate to form T and CpG steadily to TpG, which evidenced that the actually observed numbers of CpG are less than the expected (only around 21%) [25, 26]. It is the potential cement of evolution, which maybe one of the methods by which biological phenotypes by change from environmental factors are inherited.

As a DNA molecule's cytosine is methylated, there is a positional correlation between genes and DNA methylations. DNA methylations occur during or outside annotated genes in the genome (Figure 1). It is associated with gene silence that changes the biological phenotype and gene expression to determine the cellular types and functions with CpG methylation. In general, the conserved CGIs on transcription start sites (TSSs) are highly methylated and can influence transcription of genes by impeding the binding of transcriptional proteins. However, CGIs located between genes or transcriptions are observed to be highly tissue specific [27, 28], which are highly methylated and control gene expression patterns to determine the cellular types and functions in the process of cell differentiation by H3K4me3 or some potential methods [29]. The great mass of regions in the whole genome's functions is unclear. In other diseases, the region of DNA methylation change plays a role in the expression of disease-related genes and may become a new TSS. It reminds us that we should pay attention to all DNA methylation sites, not only TSS, when exploring the relation with allergy.

Machine learning is a field of computer science, which gives computer systems the ability to "lean" with data [30, 31]. Machine learning analyses data to study the construction of algorithms, which can make predictions on data, produce reliable, repeatable decisions and results, and uncover some "hidden insights" [32] and handle more complicated and bulky data. Decision tree learning is a method commonly used in data mining that includes classification tree analysis where the predicted outcome is the class to which the data belongs [33].

1.3. Idea. Whether DNA methylation such as epigenetics mark can be used to diagnose allergy and how should it be applied? It has been proposed, based on a few specific DNA methylation marks' joint detection, considering that sequencing and microarray at the whole genome are time consuming, costly, and difficult to popularize, to classify anaphylaxis types by randomForest (one of the decision tree learning) of the R programming language (R). Also, it inspires us that some researchers have presented an approach for the DNA methylation-based classification of 100 known central nervous system tumours that is based on machine learning, which can obtain accurate diagnosis and avoid observer errors by using Infinium HumanMethylation450K BeadChip arrays data recently [34]. Then, we plan to use these methods to prove the idea and find the mark types from some selected opening GEO data and the sequencing data.

2. Results

2.1. GEO Data Analysis Results. We analyzed allergy-associated hematic DNA regions' methylation from several GEO datasets by our methods. Some of the results of GSE73745 [35], GSE104471 [36], and GSE59999 [37] are shown in Figure 2. GSE73745 and GSE104471 both shared DNA regions methylation levels in monoallergen atopic asthmatic and healthy people. We discovered that these two groups obtained good classification either in t-SNE that can be seen directly (Figures 2(a) and 2(b)) or in the randomForest (the error rate: 0%) method. Then, a dataset on food allergies was found to test DNA regions' methylation levels among egg allergy patients, peanut allergy patients, and healthy people. We still observed a good classification at t-SNE (Figure 2(c)) and randomForest (Figure 2(d), the error rate: 0%). Therefore, we got a preliminary conclusion that there can be a perfect distinction between monoallergen atopic

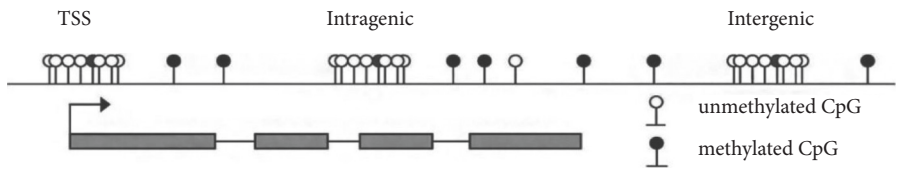


Figure 1: The ubiety between DNA methylation and gene. These include "TSS," "Intragenic," "Intergenic," and CpG alone.

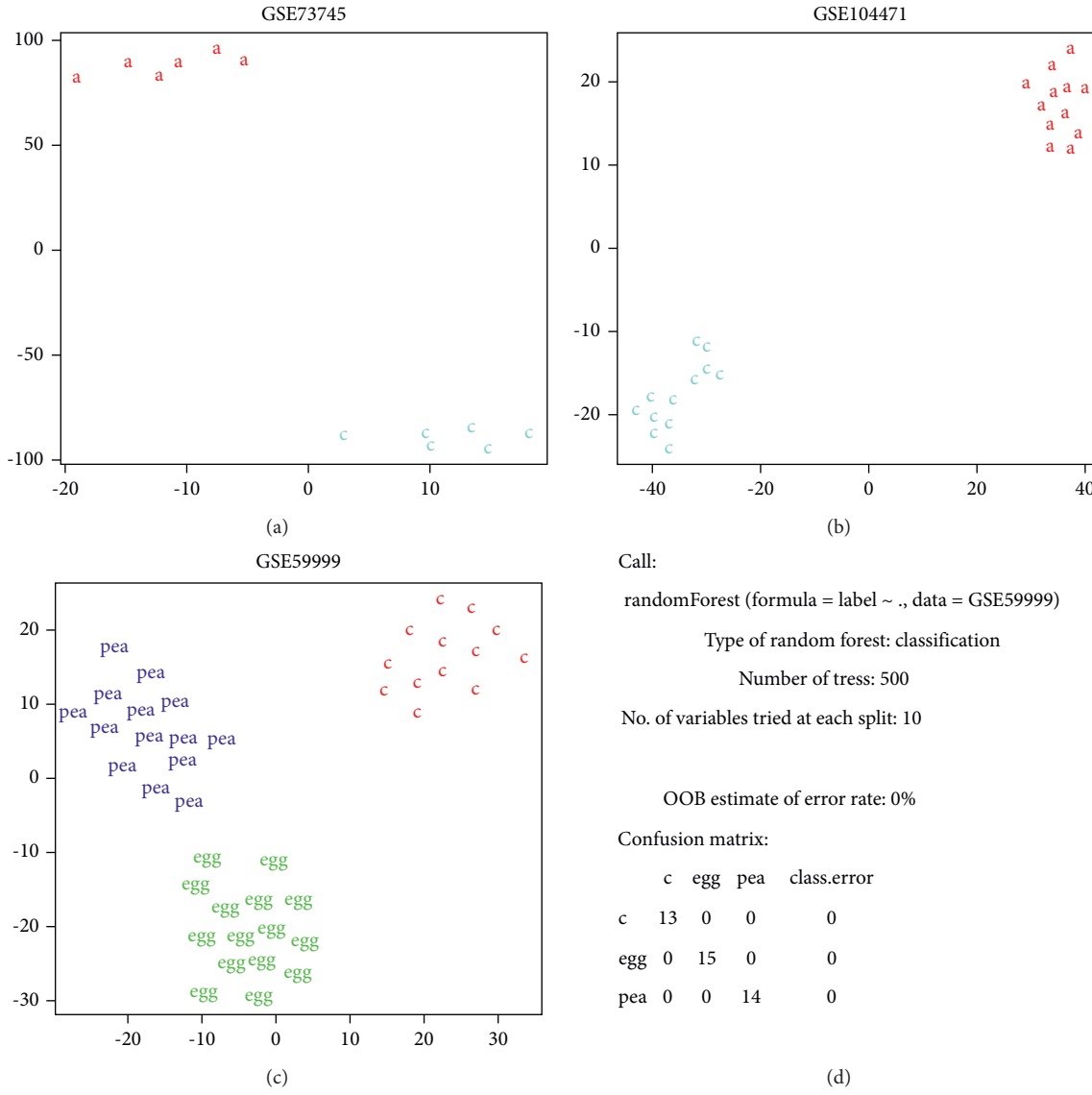


Figure 2: (a) The result of GSE73745 by t-SNE. "a" represents respiratory allergy patients, and "c" represents healthy controls. (b) The result of GSE104471 by t-SNE. "a" represents allergic asthma patients, and "c" represents healthy controls. (c) The result of GSE59999 by t-SNE. "Egg" represents egg-allergic patients, "pea" represents peanut-allergic patients, and "c" represents healthy controls. (d) The result of GSE59999 by randomForest.

asthmatics with health and those with different allergen allergies after using machine learning analysis for a few specific DNA methylation regions.

The GSE37853 [38] data described the DNA methylation levels differently in atopic allergy patients, nonatopic allergy patients, and healthy people. The GSE50222 [39] data described how the DNA methylation level changes in allergic patients correlate with symptom severity, which followed DNA methylation levels outside and during the pollen season in healthy people and allergic patients. There are

defective classifications exposed with a low error rate. Uncertain allergens and small samples available for "learning" may cause the GSE37852 data classification error (Figure 3(a)). In the GSE50222 data, allergy groups and healthy groups both recorded the same DNA methylation information at different seasons, which caused intragroup samples to become similar even if they were different and made a low error rate; however, allergic patients and healthy controls were still classified accurately (Figure 3(b)).

GSE40736 [40] data recorded allergy patients with several symptoms, and we tried to make classifications with the "Subtype" items ("non," "lung_function," "PC20," and "reversible") of this dataset. However, we failed to obtain slightly different DNA regions and had awful results (Figure 4).

2.2. Whole-Genome DNA Sequencing Results. The genome DNA sequencing data are viable after quality control (Table 1). Then, we analyzed the DNA regions (200 bp) methylation levels difference by Genome-Wide Association Studies (GWAS) and the single-nucleotide site methylation difference. The results were surprising. The GWAS result provided abundant different regions ($p < 0.05$) of annotated genes ($n = 945$), but the single-nucleotide methylation different sites had only few annotated genes ($n = 466$ of 2239), and these sites of annotated genes are always distant from the TSS. The far-TSS single-nucleotide methylation changes are always observed on dissociative single CpGs nearby some tandem repeats or intervening sequences. Also, these different single nucleotides are not detected in the different regions by GWAS.

3. Discussion

Though only few data were provided for computer-based "learning" and were not integrated (the data type: atypism; the primary data are not available), the preliminary conclusion obtained was that the specific allergen-related DNA methylation can be used to perform atopic allergen-allergy classification. However, the same method cannot discriminate the different anaphylactic symptoms. Moreover, the same method can determine whether allergic patients or healthy people or both were classified at different seasons. Therefore, we deduced that DNA methylation changed after exposure to atopic allergens, which was associated with specific allergens. These changes and anaphylactic symptoms both occurred after exposure to allergens, which may have caused the symptoms to not be classified.

In the whole process of GEO data analysis, we observed that there was a significant difference, but not high enough, and intragroup stability was low. It may be because the observed entities are DNA regions' methylation levels, which include more sequence information. In general, sequencing the genome nucleotide information, finding the difference DNA region, annotating genes or gene sites, analyzing the associated gene expression, exploring the possible regulatory mechanism, and obtaining a complete theory are a simple total process of the mechanisms associated with genetics of a disease [12, 41–44]. However, this pattern is less helpful for allergy diagnosis. But, allergen exposure can cause DNA methylation levels change, as observed in the GSE50222 data (Figure 3), and are classable. We intend to use cutting-off pilot processes by Occam's Razor, "Entities should not be multiplied unnecessarily (*Non sunt multiplicanda entia sine necessitate*)," which is the philosophy of idiographic machine learning formula selection, to find DNA nucleotides associated with methylation after allergen exposure.

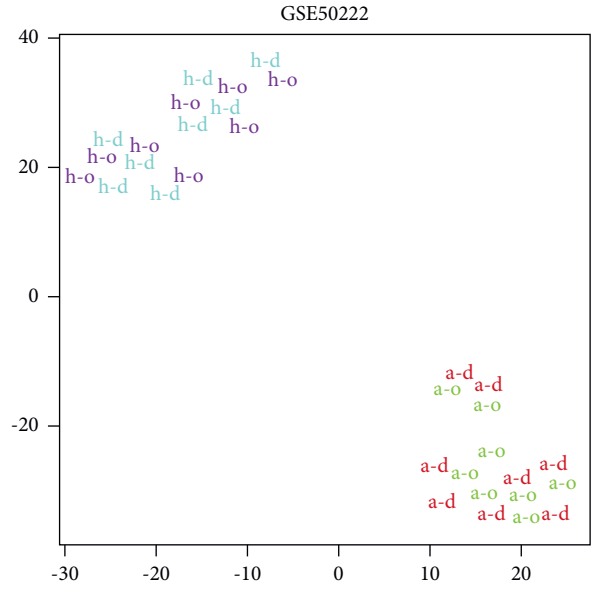
Therefore, we devised and performed experiments on peripheral blood DNA methylation sequencing to compare single-nucleotide with DNA region methylated levels, which could help to find more suitable marks. We obtained almost the opposite result for the different typological methylated levels.

The sequencing results prompt that there are some single-nucleotide methylation changes and the most changes maybe concealed from the found different DNA regions, in the nonannotated gene regions between healthy people and allergy patients after allergen exposure, in spite of the fact that the simple numbers are too low. We conjectured that DNA methylation has timely changes that should exist when allergic constitution patients are exposed to atopic allergens, and DNA methylation could be distinguished from healthy people.

Therefore, we suggest conducting more "pragmatic" research to directly find special single-nucleotide changes after exposure to an atopic allergen instead of doing complex correlativity and trying to accurately diagnose anaphylaxis using DNA methylation marks to form a single CpG methylation-based classification by machine learning. The DNA methylation potential in the field of forensic medicine is great as an epigenetic mark in spite of too little research and application. DNA methylation also captures responses to environmental factors and changes with it [14, 39], so care is needed when applying this information to personal identification. Meanwhile, our analysis results of sequencing also remind us to pay more attention to single-nucleotide methylation.

4. Methods

4.1. Data Acquisition and Processing. Searching in the Gene Expression Omnibus (GEO, https://www.ncbi.nlm.nih.gov/geo/) with "allergy," "DNA methylation," and "*Homo sapiens*" as the key words and selecting the cell type as peripheral blood mononuclear cells (PBMCs), we got 6 valuable results (GSE59999 [37], GSE73745 [35], GSE104471 [36], GSE37853 [38], GSE50222 [39], and GSE40736 [40]). In these 6 studies, the other covariates such as gender, age, and ethnicity were controlled by every independent dataset as an independent study. We have developed the same criteria for analysis to evaluate feasibility: (a) the datasets were analyzed with GEO2R [45] with $p < 0.05$ as cutoff values and extracted the β value of TOP100 probes of each sample to one document (did not include censored data); (b) we used R package Rtsne (0.15) [46] to carry out t-distributed stochastic neighbor embedding (t-SNE), which is a nonlinear dimensionality reduction technique well suited



Call:

randomForest (formula = label ~ ., data = GSE37853)

Type of random forest: classification

Number of tress: 500

No. of variables tried at each split: 7

OOB estimate of error rate: 26.67%

Confusion matrix:

	a-a	hea	n-a	class.error
a-a	3	0	1	0.2500000
hea	0	5	1	0.1666667
n-a	1	1	3	0.4000000

(a)

Call:

randomForest (formula = label ~ ., data = GSE50222)

Type of random forest: classification

Number of tress: 500

No. of variables tried at each split: 9

OOB estimate of error rate: 25%

Confusion matrix:

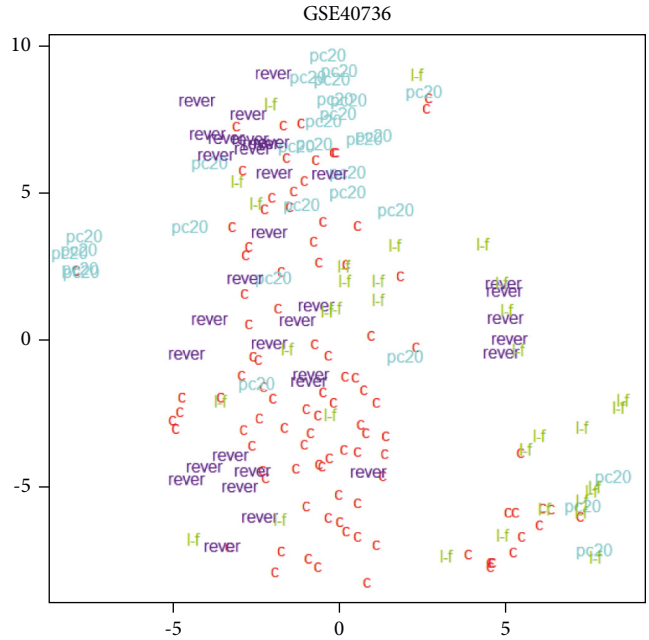
	a-d	a-o	h-d	h-o	class.error
a-d	5	3	0	0	0.375
a-o	1	7	0	0	0.125
h-d	0	0	6	2	0.250
h-o	0	0	2	6	0.250

(b)

Figure 3: (a) The result of GSE37853 by t-SNE and randomForest. "a-a" represents atopic asthmatic patients, "n-a" represents nonatopic asthmatic patients, and "hea" represents healthy controls. (b) The result of GSE50222 by t-SNE and randomForest. "a-d" represents allergic patients during the pollen season, "a-o" represents allergic patients outside of the pollen season, "h-d" represents healthy people during the pollen season, and "h-o" represents healthy people outside of the pollen season.

for embedding high-dimensional data for visualization in a low-dimensional space of two or three dimensions by setting the parameters as dims = 2, perplexity = 30(the perplexity should less than [nrow(X) − 1]/3), and max_iter = 500, to estimate the samples data visualization; (c) the data were preprocessed using R package randomForest (4.6–14) [47] by setting the parameters as tree = 500 and the training set : test set = 7 : 3 randomization to find simple classification model; and (d) R package heat-map was used to perform heat-maps, and the clustering results were added in the plots.

4.2. Study Population. We recruited participants ($n = 30$) aged between 25 and 35 who had lived in Taiyuan, Shanxi Province, for a long time and had allergic symptoms and a positive skin prick test (SPT) from September to October (wormwood pollen season). Total IgE and sIgE were assayed by the AllergyScreen® test (Mediwiss Analytic GmbH, Moers, Germany) according to the manufacturer's instructions. All the participants ($n = 30$) were tested for 19 allergens, composed of 10 types of common aeroallergens and 9 types of food allergens. Among these allergens, the



Call:

randomForest (formula = label ~ ., data = GSE40736)

Type of random forest: classification

Number of tress: 500

No. of variables tried at each split: 10

OOB estimate of error rate: 41.24%

Confusion matrix:

	c	1-f	pc20	rever	class.error
c	88	1	6	2	0.09278351
1-f	24	6	2	1	0.81818182
pc20	14	1	14	1	0.53333333
rever	21	3	4	6	0.82352941

Figure 4: The result of GSE40736 by t-SNE and randomForest. "l-f" represents lung_function, "pc20" represents PC20, "rever" represents reversible, and "c" represents healthy control.

Table 1: The sequencing quality evaluation.

Sample	Reads num.	Total bases (bp)	GC (%)	Error <1 (%)	Error <0.1 (%)
a9	305,727,176	45,859,076,400	21.34	95.31	90.30
a10	333,038,672	49,955,800,800	21.37	95.28	90.36
a12	300,768,056	45,115,208,400	21.24	95.49	90.64
c1	297,128,974	44,569,346,100	21.19	95.35	90.34
c2	301,946,840	45,292,026,000	21.23	95.23	90.25
c3	337,787,768	50,668,165,200	21.28	95.48	90.63

Error: single-nucleotide distinguished error. The quality scores all bases of per sample in the attachment document.

aeroallergens included house dust, pteronyssinus, short ragweed, estragon, mulberry, cat epithelium, dog epithelium, cockroach, mould mixture (*Penicillium notatum*, branch spore mildew, *Aspergillus fumigates,* and *Alternaria*), trees (mixture of cypress, elm, phoenix tree, *Betula*, *Fraxinus chinensis* Roxb, willow, and cottonwood), and grass (ragweed and wormwood). The food allergens included cow milk, beef, cashew-peanut-soybean, egg white/yolk, prawn, crab, cowry, mango-peach-apple-cherry, and pineapple. The sIgE level >0.35 IU/ml was considered positive, and the significant reference range of serum total IgE was defined as >100 IU/ml. Finally, only 3 participants were defined as wormwood monoallergen atopic rhinitis patients (the wormwood-specific IgE (sIgE) >17.5 IU/ml and total IgE >200 IU/ml but other common allergens sIgE all <0.35 IU/ml in peripheral blood serum). Thus, we selected the 3 wormwood monoallergen atopic rhinitis patients and 3 healthy people for methylation analysis.

4.3. Whole-Genome DNA Sequencing. Genomic DNA from peripheral blood mononuclear cell samples of wormwood monoallergen atopic rhinitis patients ($n = 3$) and healthy people ($n = 3$) during the pollen season was isolated using the DNA extraction kit (Omega, USA). We structured DNA libraries using the TruSeq Nano DNA LT Sample Prep Kit (Illumina, San Diego, CA, USA) and used bisulfite to convert all unmethylated cytosine (C) in genomic DNA into uracil (U) using EpiTect® Fast Bisulfite (Qiagen, Germany) (the bisulfite conversion rate: 99%). The DNA library quality control was performed by using the Agilent 2100 Bioanalyzer (Agilent Technologies, California). The whole-genome DNA methylation was sequenced by using a Whole-Genome Shotgun (WGS) [48] with Illumina HiSeq (Illumina, San Diego, CA, USA). After clearing the linker sequence and low-quality reads, Bismark (0.19.0) [49] was used to align the reads to the genome (GRCh38/gh38) through Bowtie2(2.2.3) [50]; then, we identified base transversion events, classified, and counted them.

4.4. Differential Methylation Analysis. RSeQC [51] (version 2.5; https://rseqc.sourceforge.net/) was used to count the distribution of methylation in different regions of genes. Differential DNA methylation regions (DMR) analysis between samples was performed on the R bioconductor

package methylKit (1.19.0) [52]. Then, we annotated location information, the chromosome segment, and upstream and downstream information by contrasting with Ensembl 89 (https://asia.ensembl.org/index.html). Following, Gene Ontology (GO, https://geneontology.org/) and Kyoto Encyclopedia of Genes and Genomes (KEGG, https://www.genome.jp/kegg/) enrichment analysis of genes was conducted to get the DMR in the promoter regions. Finally, we obtained associated different annotated genes and DNA regions. In addition, we also used the methylKit to analyze the difference between the methylation of single-nucleotide sites and annotated location information but no GO and KEGG enrichment analysis.

The analysis of the data was performed in RStudio (Version 1.2.1335; https://www.rstudio.com/) using an R environment (version 3.6.0; https://www.R-project.org). All experiments were permitted by the Ethics Committee of Shanxi Medical University (2017LL073), and all methods were performed in accordance with the relevant guidelines and regulations.

Data Availability

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Conflicts of Interest

The authors declare no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Authors' Contributions

Xiangjie Guo, Yaqin Bai, and Hualin Guo contributed equally to this work.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (No. 81971790), Natural Science Foundation of Shanxi Province (201601D102070), and Shanxi Medical University foundation for doctors (03201534).

References

[1] Y. Wang, K. J. Allen, N. H. A. Suaini, V. McWilliam, R. L. Peters, and J. J. Koplin, "The global incidence and prevalence of anaphylaxis in children in the general population: a systematic review," *Allergy*, vol. 74, no. 6, pp. 1063–1080, 2019.

[2] T. C. Yao, A. C. Wu, Y. W. Huang, J. Y. Wang, and H. J. Tsai, "Increasing trends of anaphylaxis-related events: an analysis of anaphylaxis using nationwide data in Taiwan, 2001–2013," *World Allergy Organization Journal*, vol. 11, p. 23, 2018.

[3] W. E. Berger, "Overview of allergic rhinitis," *Annals of Allergy, Asthma & Immunology*, vol. 90, no. 6, pp. 7–12, 2003.

[4] J. L. Brożek, J. Bousquet, I. Agache et al., "Allergic rhinitis and its impact on asthma (ARIA) guidelines-2016 revision," *The Journal of Allergy and Clinical Immunology*, vol. 140, no. 4, pp. 950–958, 2017.

[5] M. L. K. Tang and R. J. Mullins, "Food allergy: is prevalence increasing?" *Internal Medicine Journal*, vol. 47, no. 3, pp. 256–261, 2017.

[6] J. J. Koplin, E. N. C. Mills, and K. J. Allen, "Epidemiology of food allergy and food-induced anaphylaxis: is there really a Western world epidemic?" *Current Opinion in Allergy & Clinical Immunology*, vol. 15, no. 5, pp. 409–416, 2015.

[7] P. Giavina-Bianchi, M. V. Aun, and J. Kalil, "Drug-induced anaphylaxis: is it an epidemic?" *Current Opinion in Allergy & Clinical Immunology*, vol. 18, no. 1, pp. 59–65, 2018.

[8] S. Lee, E. P. Hess, C. Lohse, W. Gilani, A. M. Chamberlain, and R. L. Campbell, "Trends, characteristics, and incidence of anaphylaxis in 2001–2010: a population-based study," *Journal of Allergy and Clinical Immunology*, vol. 139, no. 1, pp. 182–188, 2017.

[9] C. Flohr and J. Mann, "New insights into the epidemiology of childhood atopic dermatitis," *Allergy*, vol. 69, no. 1, pp. 3–16, 2014.

[10] D. Vercelli, "Discovering susceptibility genes for asthma and allergy," *Nature Reviews Immunology*, vol. 8, no. 3, pp. 169–182, 2008.

[11] G. A. Lockett and J. W. Holloway, "Genome-wide association studies in asthma; perhaps, the end of the beginning," *Current Opinion in Allergy & Clinical Immunology*, vol. 13, no. 5, pp. 463–469, 2013.

[12] K. Bønnelykke, M. C. Matheson, T. H. Pers et al., "Meta-analysis of genome-wide association studies identifies ten loci influencing allergic sensitization," *Nature Genetics*, vol. 45, no. 8, pp. 902–906, 2013.

[13] M. F. Moffatt and W. O. Cookson, "The genetics of asthma. Maternal effects in atopic disease," *Clinical and Experimental Allergy: Journal of the British Society for Allergy and Clinical Immunology*, vol. 28, no. Suppl 1, pp. 56–61, 1998.

[14] L. Liang, S. A. G. Willis-Owen, C. Laprise et al., "An epigenome-wide association study of total serum immunoglobulin E concentration," *Nature*, vol. 520, no. 7549, pp. 670–674, 2015.

[15] A. P. Feinberg and B. Tycko, "The history of cancer epigenetics," *Nature Reviews Cancer*, vol. 4, no. 2, pp. 143–153, 2004.

[16] M. D. Anway, A. S. Cupp, M. Uzumcu, and M. K. Skinner, "Epigenetic transgenerational actions of endocrine disruptors and male fertility," *Science*, vol. 308, no. 5727, pp. 1466–1469, 2005.

[17] A.-K. M. Sjogren, F. Barrenas, A. Muraro et al., "Monozygotic twins discordant for intermittent allergic rhinitis differ in mRNA and protein levels," *Allergy*, vol. 67, no. 6, pp. 831–833, 2012.

[18] X. Liu, S. Zhang, H. J. Tsai et al., "Genetic and environmental contributions to allergen sensitization in a Chinese twin study," *Clinical & Experimental Allergy*, vol. 39, no. 7, pp. 991–998, 2009.

[19] M. Pascual, M. Suzuki, M. Isidoro-Garcia et al., "Epigenetic changes in B lymphocytes associated with house dust mite allergic asthma," *Epigenetics*, vol. 6, no. 9, pp. 1131–1137, 2011.

[20] C. V. Breton, H. M. Byun, M. Wenten, F. Pan, A. Yang, and F. D. Gilliland, "Prenatal tobacco smoke exposure affects global and gene-specific DNA methylation," *American Journal of Respiratory and Critical Care Medicine*, vol. 180, no. 5, pp. 462–467, 2009.

[21] E. Morales, M. Bustamante, N. Vilahur et al., "DNA hypomethylation at ALOX12 is associated with persistent wheezing

in childhood," *American Journal of Respiratory and Critical Care Medicine*, vol. 185, no. 9, pp. 937–943, 2012.

[22] A. P. Bird, "CpG-rich islands and the function of DNA methylation," *Nature*, vol. 321, no. 6067, pp. 209–213, 1986.

[23] M. G. Guenther, S. S. Levine, L. A. Boyer, R. Jaenisch, and R. A. Young, "A chromatin landmark and transcription initiation at most promoters in human cells," *Cell*, vol. 130, no. 1, pp. 77–88, 2007.

[24] D. M. Jeziorska, R. J. S. Murray, M. De Gobbi et al., "DNA methylation of intragenic CpG islands depends on their transcriptional activity during differentiation and disease," *Proceedings of the National Academy of Sciences*, vol. 114, no. 36, pp. E7526–e7535, 2017.

[25] E. S. Lander, L. M. Linton, B. Birren et al., "Initial sequencing and analysis of the human genome," *Nature*, vol. 409, no. 6822, pp. 860–921, 2001.

[26] International Human Genome Sequencing Consortium, "Finishing the euchromatic sequence of the human genome," *Nature*, vol. 431, pp. 931–945, 2004.

[27] A. K. Maunakea, R. P. Nagarajan, M. Bilenky et al., "Conserved role of intragenic DNA methylation in regulating alternative promoters," *Nature*, vol. 466, no. 7303, pp. 253–257, 2010.

[28] R. Illingworth, A. Kerr, D. DeSousa et al., "A novel CpG island set identifies tissue-specific methylation at developmental gene loci," *PLoS Biology*, vol. 6, no. 1, p. e22, 2008.

[29] A. M. Deaton, S. Webb, A. R. W. Kerr et al., "Cell type-specific DNA methylation at intragenic CpG islands in the immune system," *Genome Research*, vol. 21, no. 7, pp. 1074–1086, 2011.

[30] A. L. Samuel, "Some studies in machine learning using the game of checkers," *IBM Journal of Research and Development*, vol. 3, no. 3, pp. 210–229, 1959.

[31] J. R. Koza, F. H. Bennett, D. Andre, and M. A. Keane, *Artificial Intelligence in Design '96*, J. S. Gero and S. Fay, Eds., Springer, Berlin, Germany, 1996.

[32] R. P. Hall, B. Falkenhainer, N. Flann et al., "A review of the fourth international workshop on machine learning," *Machine Learning*, vol. 2, no. 2, pp. 173–190, 1987.

[33] P. E. Utgoff, N. C. Berkman, and J. A. Clouse, "Decision tree induction based on efficient tree restructuring," *Machine Learning*, vol. 29, no. 1, pp. 5–44, 1997.

[34] D. Capper, D. T. W. Jones, M. Sill et al., "DNA methylation-based classification of central nervous system tumours," *Nature*, vol. 555, no. 7697, pp. 469–474, 2018.

[35] S. A. S. Langie, K. Szarc vel Szic, K. Declerck et al., "Whole-genome saliva and blood DNA methylation profiling in individuals with a respiratory allergy," *PLoS One*, vol. 11, no. 3, Article ID e0151109, 2016.

[36] I. V. Yang, A. Richards, E. J. Davidson et al., "The nasal methylome: a key to understanding allergic asthma," *American Journal of Respiratory and Critical Care Medicine*, vol. 195, no. 6, pp. 829–831, 2017.

[37] D. Martino, T. Dang, A. Sexton-Oates et al., "Blood DNA methylation biomarkers predict clinical reactivity in food-sensitized infants," *Journal of Allergy and Clinical Immunology*, vol. 135, no. 5, pp. 1319–1328, 2015.

[38] D. Stefanowicz, T. L. Hackett, F. S. Garmaroudi et al., "DNA methylation profiles of airway epithelial cells and PBMCs from healthy, atopic and asthmatic children," *PLoS One*, vol. 7, no. 9, Article ID e44213, 2012.

[39] C. E. Nestor, F. Barrenäs, H. Wang et al., "DNA methylation changes separate allergic patients from healthy controls and may reflect altered CD4+ T-cell population structure," *PLoS Genetics*, vol. 10, no. 1, Article ID e1004059, 2014.

[40] I. V. Yang, B. S. Pedersen, A. Liu et al., "DNA methylation and childhood asthma in the inner city," *Journal of Allergy and Clinical Immunology*, vol. 136, no. 1, pp. 69–80, 2015.

[41] L. Machado, J. Esteves de Lima, O. Fabre et al., "In situ fixation redefines quiescence and early activation of skeletal muscle stem cells," *Cell Reports*, vol. 21, no. 7, pp. 1982–1993, 2017.

[42] A. B. Hart and H. R. Kranzler, "Alcohol dependence genetics: lessons learned from genome-wide association studies (GWAS) and post-GWAS analyses," *Alcoholism: Clinical and Experimental Research*, vol. 39, no. 8, pp. 1312–1327, 2015.

[43] S. C. Mack and P. A. Northcott, "Genomic analysis of childhood brain tumors: methods for genome-wide discovery and precision medicine become mainstream," *Journal of Clinical Oncology*, vol. 35, no. 21, pp. 2346–2354, 2017.

[44] S. S. Fei, A. D. Mitchell, M. B. Heskett et al., "Patient-specific factors influence somatic variation patterns in von hippel-lindau disease renal tumours," *Nature Communications*, vol. 7, no. 1, Article ID 11588, 2016.

[45] T. Barrett, S. E. Wilhite, P. Ledoux et al., "NCBI GEO: archive for functional genomics data sets--update," *Nucleic Acids Research*, vol. 41, no. D1, pp. D991–D995, 2013.

[46] G. H. Laurens van der Maaten, "Visualizing data using t-SNE," *Journal of Machine Learning Research*, vol. 9, pp. 2579–2605, 2008.

[47] L. Breiman and A. Cutler, *randomForest: Breiman and Cutler's Random Forests for Classification and Regression*, https://cran.r-project.org/web/packages/randomForest/randomForest.pdf, 2008.

[48] J. L. Weber and E. W. Myers, "Human whole-genome shotgun sequencing," *Genome Research*, vol. 7, no. 5, pp. 401–409, 1997.

[49] F. Krueger and S. R. Andrews, "Bismark: a flexible aligner and methylation caller for bisulfite-seq applications," *Bioinformatics*, vol. 27, no. 11, pp. 1571-1572, 2011.

[50] B. Langmead and S. L. Salzberg, "Fast gapped-read alignment with bowtie 2," *Nature Methods*, vol. 9, no. 4, pp. 357–359, 2012.

[51] L. Wang, S. Wang, and W. Li, "RSeQC: quality control of RNA-seq experiments," *Bioinformatics*, vol. 28, no. 16, pp. 2184-2185, 2012.

[52] A. Akalin, M. Kormaksson, S. Li et al., "methylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles," *Genome Biology*, vol. 13, no. 10, p. R87, 2012.