

## Article

# Environmental Factor Index (EFI): A Novel Approach to Measure the Strength of Environmental Influence on DNA Methylation in Identical Twins

Yoichi Takenaka <sup>1,2,\*</sup>, Osaka Twin Research Group <sup>2,†</sup> and Mikio Watanabe <sup>2,3</sup><sup>1</sup> Faculty of Informatics, Kansai University, Osaka 569-1052, Japan<sup>2</sup> Center for Twin Research, Graduate School of Medicine, The University of Osaka, Osaka 565-0871, Japan; nabe@sahs.med.osaka-u.ac.jp (M.W.)<sup>3</sup> Department of Clinical Laboratory and Biomedical Sciences, Graduate School of Medicine, The University of Osaka, Osaka 565-0871, Japan

\* Correspondence: takenaka@kansai-u.ac.jp

† Osaka Twin Research Group: Norio Sakai, Masanori Takahashi, Teiji Nishio, Kei Kamide, Shinji Kihara, Hiroko Watanabe, Mikio Watanabe and Dousatsu Sakata.

**Abstract:** Background/Objectives: The dynamic interaction between genomic DNA, epigenetic modifications, and phenotypic traits was examined in identical twins. Environmental perturbations can induce epigenetic changes in DNA methylation, influencing gene expression and phenotypes. Although DNA methylation mediates gene-environment correlations, the quantitative effects of external factors on DNA methylation remain underexplored. This study aimed to quantify these effects using a novel approach. Methods: A cohort study was conducted on healthy monozygotic twins to evaluate the influence of environmental stimuli on DNA methylation. We developed the Environmental Factor Index (EFI) to identify methylation sites showing statistically significant changes in response to environmental stimuli. We analyzed the identified sites for associations with disorders, DNA methylation markers, and CpG islands. Results: The EFI identified methylation sites that exhibited significant associations with genes linked to various disorders, particularly cancer. These sites were overrepresented on CpG islands compared to other genomic features, highlighting their regulatory importance. Conclusions: The EFI is a valuable tool for understanding the molecular mechanisms underlying disease pathogenesis. It provides insights into the development of preventive and therapeutic strategies and offers a new perspective on the role of environmental factors in epigenetic regulation.

**Keywords:** epigenetics; DNA methylation; environmental factors; twin study



**Citation:** Takenaka, Y.; Osaka Twin Research Group; Watanabe, M. Environmental Factor Index (EFI): A Novel Approach to Measure the Strength of Environmental Influence on DNA Methylation in Identical Twins. *Epigenomes* **2024**, *8*, 44. <https://doi.org/10.3390/epigenomes8040044>

Academic Editors: Ivana De la Serna and Che-Kun James Shen

Received: 3 September 2024

Revised: 13 November 2024

Accepted: 19 November 2024

Published: 21 November 2024



**Copyright:** © 2024 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 (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

DNA methylation is an important epigenetic component. Although all of the cells in multicellular organisms have identical genomes, differentiated cells, such as those in the liver and muscle, appear different, with varied roles; however, the phenotypes at the cellular level may differ within the same genome (DNA sequence). This concept is known as epigenetics and is defined as a “stably heritable phenotype resulting from changes in a chromosome without alterations in the DNA sequence” [1].

DNA methylation is the process in which methyl groups are added primarily to cytosine residues of genomic DNA [2–6]. The methylation state of the genome is inherited from cell division [7,8] and plays an important role in normal development and cell differentiation in vertebrates [9,10]. Interestingly, DNA methylation alters gene expression, and methylation occurring in the promoter regions inhibits DNA binding to transcription factors [11], thus decreasing their expression. Conversely, the methylation of gene bodies promotes transcription [12,13]. These phenomena maintain the tissue- or cell-type-specific function of differentiated cells. Therefore, it can be inferred that DNA methylation stably

alters gene expression profiles to remember its location/position in the cell differentiation tree [14].

DNA methylation was recently identified as the core of the etiology of human disorders. The dysregulation of genes resulting from DNA methylation causes various neurodevelopmental syndromes, such as fragile X and Rett, immunodeficiency, centromeric region instability, and facial anomalies (ICF) [15]. Abnormal DNA methylation is also associated with breast [16,17], cervical [16,18], and other cancers [19–24], as well as diabetes [25–27], kidney disease [28,29], autoimmune disorders [30–32], and allergies [33,34]. Moreover, the rate of cell aging can be quantified based on DNA methylation levels [35–39].

Because the correlation between a single DNA methylation site and the resulting phenotype is low, many DNA methylation sites are associated with various phenotypes. For example, the number of sites needed to measure cell age was 353 for the Horvath clock [35,39], 71 for the Hannum epigenetic age [36], and 513 for the Levine clock [38]. These numerous methylation sites resulted from an abundance of positions with methylation potential. The human genome contains >20 million CpG dinucleotide sequences that are methylation candidates. Therefore, machine learning has proven to be a valuable tool to determine these associations.

Genetic and environmental factors can have an impact on DNA methylation, which in turn, affects the phenotype. Genetic factors are determined solely by inheritance, whereas environmental factors include climate, nutrition, lifestyle, stress, drugs, and living conditions. The combined influence of genetic and environmental factors can be observed in various phenotypic outcomes. For example, monogenic diseases arise from mutations in a single gene, whereas dietary habits are considered one of the causes of colorectal cancer [40]. This effect on the initiation and progression of various diseases has also been demonstrated through insurance claims [41] and studies on twins [42].

DNA methylation is an important mechanism that connects genetic and environmental factors to a particular phenotype. For example, exposure to chemicals, such as polychlorinated biphenyls, lead, and bisphenol A, has been associated with an increased risk of autism spectrum disorders, potentially through alterations in DNA methylation [43]. In addition, children can inherit parental stress [44] and nutritional status [45] through DNA methylation. Epigenomic changes induced by environmental factors also influence the development of more diseases compared with genetic elements [46]. Thus, determining the effects of genetic and environmental factors on DNA methylation will improve our understanding of the etiology and pathogenesis of these diseases [47].

Should the contributions of environmental and genetic factors to DNA methylation be successfully separated, this could significantly aid in unraveling the causes of diseases and developing novel therapeutic approaches. Fujii found that DNA methylation is a mediator of associations between environment and diseases [48]. However, the interplay between genetic and environmental factors in DNA methylation denotes a complex mechanism, making separating their respective contributions challenging from biological and technical perspectives [49,50] except when using monozygotic twins.

Monozygotic twins share the same genetic material and are used to differentiate the effects of genetic and environmental factors on DNA methylation levels [51,52]. Based on this unique feature, twin studies enhance our understanding of the processes regulating epigenetic variation and unravel the relative contributions of environmental and genetic factors to complex traits [53,54]. Building upon these findings, we proposed an index to measure the influence of environmental factors on each methylation site. This index compares the differences in methylation levels between younger twins and elderly twins. Overall, our analysis provides insight into the complex interplay between DNA methylation and environmental factors that have implications for disease development and prevention.

To address the gap in the understanding of the influence of environmental stimuli on DNA methylation, we introduced the Environmental Factor Index (EFI) in this study. We designed the EFI to identify the methylation sites that show statistically significant changes in response to environmental stimuli. By comparing the methylation patterns

between younger and elderly twin pairs, we aim to quantify the influence of environmental factors on specific methylation sites. This approach provides insights into the potential regulatory mechanisms and the association of methylation changes with various disorders, particularly those linked to CpG islands. The EFI represents a novel tool for advancing our understanding of the molecular mechanisms underlying disease pathogenesis and developing preventive and therapeutic strategies.

Current methods, such as a correlation analysis between methylation levels and age and epigenome-wide association studies (EWAS) to assess environmental impacts for analyzing the impact of environmental factors on DNA methylation often fail to effectively differentiate the influence of genetic factors from environmental stimuli. Traditional approaches may lack the statistical power or specificity to identify subtle yet significant epigenetic changes linked to environmental factors. We explicitly designed the EFI to address these challenges by utilizing monozygotic twins, thereby more precisely isolating the impact of environmental factors.

Unlike previous metrics, the EFI provides a novel approach that focuses on isolating the specific environmental contributions to methylation changes, which is particularly crucial in understanding disease mechanisms. By comparing younger and older twin pairs, the EFI offers a unique lens through which we can observe the cumulative impact of environmental exposure over time, something that previous studies have struggled to accurately quantify.

## 2. Results

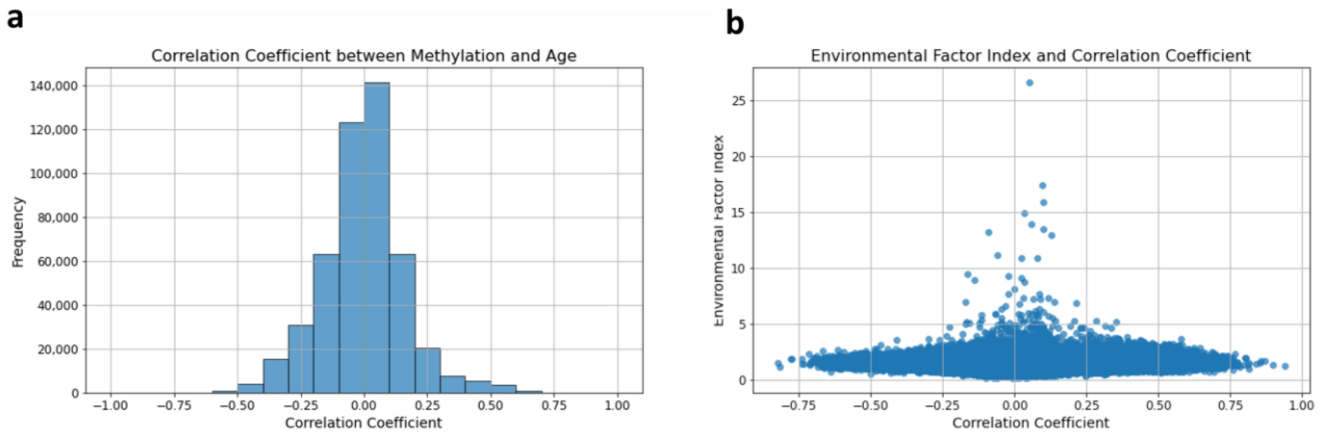
### 2.1. Difference Between EFI and Correlation Coefficient

The EFI, a method that intricately divides twin pairs into two groups based on their age, presents a novel approach that addresses the complexity of environmental influences on DNA methylation. Conversely, the most direct method to explore the connection between DNA methylation and age is to evaluate the linear correlation between methylation levels and chronological age. The EFI was calculated for each probe, and its values are derived using all twin pairs. Therefore, more than a direct comparison between the EFI and chronological age is needed. To address this, we calculated each probe's correlation coefficient between age and methylation levels and compared these coefficients with the EFI values.

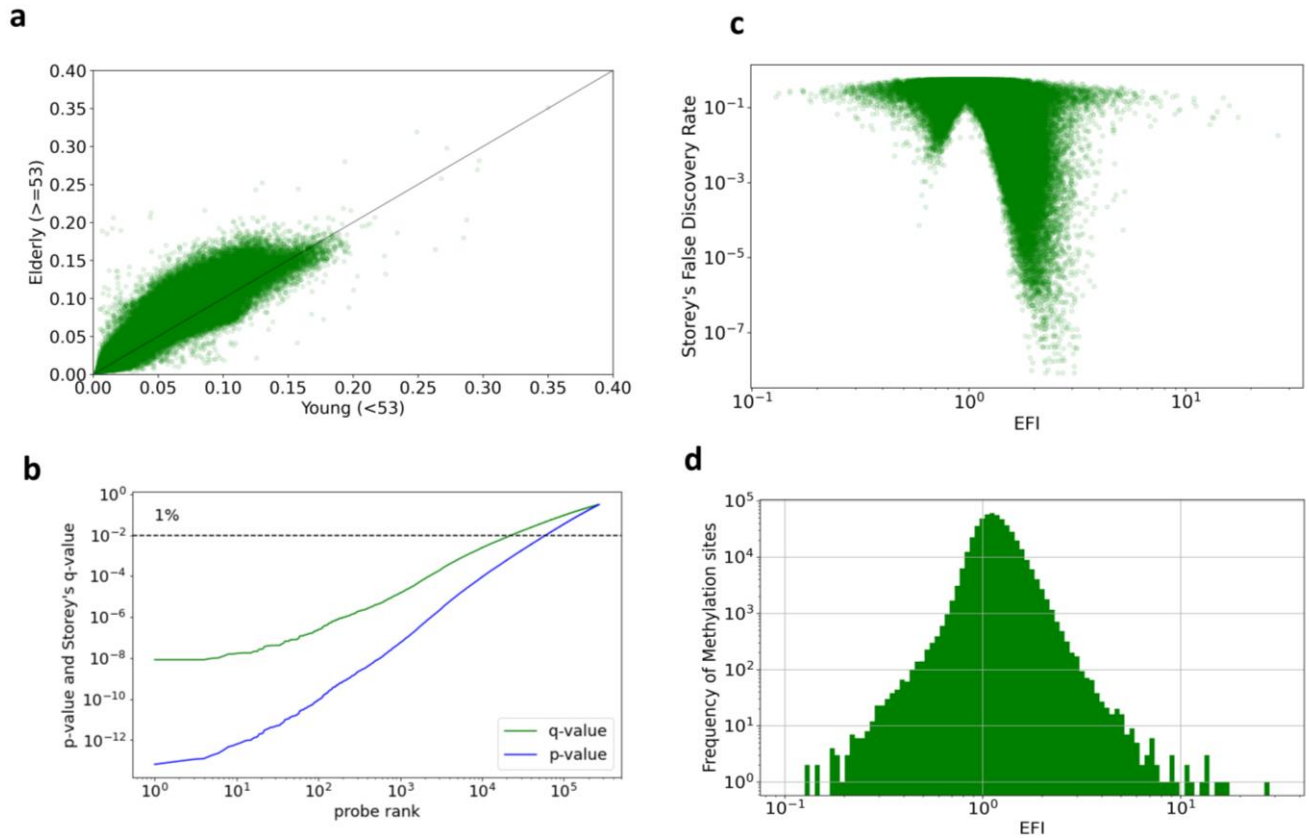
Figure 1a presents the distribution of correlation coefficients between DNA methylation levels and chronological age for each DNA methylation site. The distribution of these correlation coefficients is unimodal, indicating a single peak. This characteristic is similar to the unimodal distribution observed in Figure 2d for the EFI values.

Figure 1b plots the EFI values against the correlation coefficients for each methylation site. Each point on the scatter plot represents a methylation site, with the EFI value on the  $y$ -axis and the corresponding correlation coefficient on the  $x$ -axis. The lack of a clear relationship between EFI values and correlation coefficients indicates that the EFI captures aspects of the methylation data that are only partially dependent on linear age–methylation relationships.

The differences in DNA methylation levels between twin pairs are primarily explained by the influence of environmental factors. If the influence of environmental factors is neutral, it can generally be assumed that the longer the exposure to the environmental factors (i.e., the older the individual), the greater the differences in DNA methylation levels. This hypothesis is supported by Figure 2a, where 76.5% of the DNA methylation sites show larger differences in the elderly group compared to the younger group.



**Figure 1.** Difference between the EFI and coefficient correlation of DNA methylation values and age: (a) distribution of correlation coefficients between methylation values and chronological age for each DNA methylation site; (b) scatter plot of EFI and the correlation coefficient. Each point represents a methylation site, plotting its EFI value on the *y*-axis against its corresponding correlation coefficient on the *x*-axis.



**Figure 2.** Statistics on EFI. (a) Scatter chart of SDs for young twins (<53 years-old) and elderly twins ( $\geq 53$  years-old). The *x*- and *y*-axes correspond to the denominator and numerator of the EFI, respectively. Each dot represents a methylation site. (b) *p*-values of Levene's test and *q*-values representing the probabilities adjusted for multiple testing correction. (c) Scatter chart of EFI and FDR from Levene's test between young and elderly twins. Each dot represents a methylation site. Both axes are logarithmic. (d) Histogram of EFI. Both axes are logarithmic.

If the EFI were strongly correlated with age, we would expect a trend where higher correlation coefficients correspond to larger EFI values. However, such a clear trend is not

observed in Figure 1b, suggesting that the EFI captures non-linear aspects of age-related methylation changes or other environmental influences beyond simple chronological age.

These results underscore the idea that the EFI and correlation coefficients are fundamentally different metrics. While correlation coefficients measure the strength of the linear relationship between age and methylation at each site, the EFI, with its unique approach of using twin pairs to eliminate the influence of genetic factors, presents a significant distinction from values derived solely from the relationship with age, such as correlation coefficients. This unique aspect of the EFI provides new insights that may be of interest to the scientific community.

## 2.2. Methylation Sites Are Linked to Disorders

The current study was conducted on identical twins to determine the effect of environmental factors on DNA methylation at each site. The results indicated that the degree of methylation increased at 22,568 sites in an age-dependent manner, whereas a decrease was observed at 94 sites. We examined the correlation among significant methylation sites, disease, and CpG.

As detailed in the Introduction, DNA methylation has been reported to be associated with various diseases. Therefore, we examined the link between the significant methylation sites and such disorders. The top 10 sites were identified (Table 1) along with the probe ID on the Infinium HumanMethylation450 BeadChip Kit, the gene symbol of the site, the EFI, and the related disorders. We linked the association between these sites and diseases based on the relationship between the gene symbols of the sites and elite genes from MalaCards [55], which is a database of human diseases (accessed 10 June 2021).

**Table 1.** Top ten methylation sites in the EFI.

Probe ID	Gene Symbol	EFI	q-Value	Disorder
cg11539424	<i>CLGN</i>	5.22	0.057%	
cg25105066	<i>AUTS2</i>	5.14	0.0055%	Intellectual Developmental Disorder, Autosomal Dominant 26
cg14464244	<i>MAGI2</i>	4.76	0.047%	Nephrotic Syndrome, Type 15 Genetic Steroid-Resistant Nephrotic Syndrome
cg06445586		4.71	0.31%	
cg02878907	<i>ZNF709</i>	4.65	0.034%	
cg04883656	<i>OGFRL1</i>	4.34	0.023%	
cg21155461	<i>ZNF544</i>	4.04	0.68%	
cg21364278		4.01	0.37%	
cg17289202	<i>ZNF532</i>	3.95	0.18%	
cg15368722		3.86	0.084%	

Two methylation sites have been linked to disorders. One of these, known as cg25105066, belongs to the autism susceptibility candidate 2 (*AUTS2*) gene, which is an “Activator of Transcription and Developmental Regulator *AUTS2*” and an elite gene for “intellectual developmental disorder autosomal dominant 26”. A previous study found that DNA methylation of *AUTS2* is linked to this disorder [56]. Another study discovered that DNA methylation of *AUTS2* in the placenta is associated with neurodevelopment in children [57]. Thus, *AUTS2* may be used as a biomarker for autism spectrum disorder risk. The other methylation site, cg14464244, belongs to the *MAGI2* gene (i.e., “Membrane Associated Guanylate Kinase, WW and PDZ Domain Containing 2”). DNA methylation of *MAGI2* has been linked to various cancers [58–60], whereas no correlation was established with a nephrotic syndrome.

Table 2 presents the link between disorders and genes strongly influenced by environmental factors. This Table lists the top ten genes with a high number of significant sites.



The columns contain the gene symbols, the number of sites, and the associated disorders. We used the elite genes, as defined by MalaCards, to establish the association between genes and disorders (accessed on 10 June 2021).

**Table 2.** Top ten genes in the EFI.

Gene Symbol	#Significant Site	Disorder
<i>PTPRN2</i>	39	
<i>TNXB</i>	34	Ehlers–Danlos Syndrome, Classic-Like Vesicoureteral Reflux 8
<i>PRDM16</i>	31	Left Ventricular Noncompaction 8
<i>BRUNOL4</i>	28	
<i>COL11A2</i>	25	Otospondylomegaepiphyseal Dysplasia, Autosomal Dominant/Recessive
<i>NKX6-2</i>	24	Spastic Ataxia 8, Autosomal Recessive, with Hypomyelinating Leukodystrophy
<i>PCDHGA4</i>	21	
<i>THRB</i>	20	Thyroid Hormone Resistance, Generalized, Autosomal Dominant Thyroid Hormone Resistance, Selective Pituitary
<i>MAGI2</i>	20	Nephrotic Syndrome, Type 15 Genetic Steroid-Resistant Nephrotic Syndrome
<i>TP73</i>	20	Small Cell Cancer of the Lung

Table 3 summarizes the findings from Tables 1 and 2. We identified two sites and seven genes linked to disorders using the elite genes of MalaCards. This Table shows that of the top ten methylation sites selected by EFI, seven were annotated as genes, and among those, two were classified as elite genes. Similarly, seven of the top 10 genes were classified as elite. Based on this, the elite gene ratio for the top ten sites was calculated as  $2/7 = 29\%$ , and for the top ten genes, the elite gene ratio was  $7/10 = 70\%$ .

**Table 3.** Statistical tests for predominantly high levels of disease-related genes.

	Top 10 Sites	Top 10 Genes	Random
#Sites	10	-	100
#Genes	7	10	77
#Elite genes	2	7	19
Elite gene ratio	29%	70%	25%
<i>p</i> -value	1.000	0.0067	-
Odds ratio	0.86	7.12	-

As detailed in Section 4.5, we performed Fisher’s exact test, and for the top 10 genes, the null hypothesis was rejected, indicating that the association between the elite genes and disorders was statistically significant.

This analysis suggests that a single site with variable methylation levels resulting from environmental factors is unlikely to be directly related to a disorder. However, when methylation changes are clustered within a single gene, the likelihood of an association with the disorder increases compared with chance, highlighting the potential relevance of our research. Nevertheless, it is essential to acknowledge that methylation changes may also result from the disorder rather than the cause.

### 2.3. Environmental Factors Alter DNA Methylation Levels in Methylation Markers

Genetic markers have been used to detect disorders and can be classified into three categories: (1) biochemical markers are in the blood or other body fluids, indicating the presence of disorders; (2) molecular markers which indicate specific changes

or alterations in DNA sequences associated with disease; and (3) methylation markers indicating specific changes in DNA methylation patterns associated with disorders. These markers may be used as diagnostic or prognostic tools as well as potential targets for therapeutic interventions [61,62]. As the methylation levels of these markers alter in response to the environment, they be significant sites for the EFI to accurately capture the environmental factors.

Table 4 lists the number of methylation markers identified for colorectal [19–21,63], breast [16,17], cervical [16,18], and lung [22–24] cancers, along with the number of genes associated with significant sites. The columns list the total number of biomarkers, the number of genes with significant sites, and the *p*-value obtained from the binomial test. A list of marker genes for each cancer and the number of significant sites is available in the Supporting Information file: Marker.pdf.

**Table 4.** Statistical tests for predominantly high levels of disease-related genes.

	#Markers	#Significant Sites	Ratio	<i>p</i> -Value
Colorectal	51	39	76%	$3.80 \times 10^{-11}$
Breast	11	11	100%	$2.71 \times 10^{-6}$
Cervical	7	7	100%	$2.87 \times 10^{-4}$
Lung	16	15	94%	$2.91 \times 10^{-7}$
Total	85	72	85%	$2.11 \times 10^{-24}$

A large proportion (85%) of the markers consisted of significant genes, which was higher than the probability of this occurring by chance (31.2%). The results indicated that the EFI can assess the impact of environmental factors on various cancers. The strength of the environmental factors may be evaluated by comparing the variations in the methylation levels between twins.

#### 2.4. DNA Methylation on CpG Islands

CpG islands are genome regions containing a high frequency of CpG sites. Approximately 70% of the proximal promoters in humans located near the transcription start site contain CpG islands [64]. Typically, the methylation of CpG islands is associated with transcriptional repression, long-term gene silencing, X-chromosome inactivation, genomic imprinting, and pre-mRNA alternative splicing [65,66]. Recent studies have indicated a regulatory role for DNA methylation [67] and that the methylation of CpG sites in CpG islands alters gene expression [68].

Table 5 lists the information on five different features of CpG methylation, including the total number of methylation sites analyzed using the BeadChip, the number of significant sites, and the ratio of significant sites to the total number of sites. The CpG features were defined as follows: CpG islands include regions >500 bp, >55% GC, and an expected/observed CpG ratio of >0.65. Of note, 40% of the gene promoters contain islands [69], whereas shores are regions located 0–2 kbp from CpG islands and consist of >75% of tissue-specific differentially methylated regions. The methylation in shores is more strongly correlated to gene expression compared with that of the CpG islands [70,71]. Shelves are the 2–4 kbp regions from the islands. North and south indicate upstream and downstream to the CpG island, respectively.

The distribution in Table 5 shows significant methylation sites across different genomic features. This Table shows the number of sites (#sites) and significant sites (#significant sites) for each genomic feature, including North Shelf, North Shore, CpG Islands, South Shore, South Shelf, and Others. The column “% of significant sites in Feature” represents the proportion of significant sites within each feature, calculated by dividing the number of significant sites by the total number of sites in that feature. The column “% of significant sites in total” indicates the proportion of significant sites in the entire dataset, calculated

by dividing the number of significant sites in each feature by the total number of probes (481,190). These values allow for a comparison of the distribution of significant sites across various genomic regions.

**Table 5.** Significant sites for each CpG feature.

Feature	#Sites	#Significant Sites	% of Significant Sites	
			in Feature	in Total
North Shelf	24,716	801	3.2%	0.17%
North Shore	62,647	2422	3.9%	0.50%
CpG Island	149,935	10,898	7.3%	2.3%
South Shore	49,055	1850	3.8%	0.38%
South Shelf	22,182	712	3.2%	0.15%
Other	172,655	6069	3.5%	1.3%
Total	481,190	22,752	4.7%	4.7%

The proportion of significant sites in the CpG islands was greater than that of the other features, suggesting that CpG islands are highly responsive to environmental factors. Because of their enrichment in regulatory regions, association with tissue-specific gene expression, and susceptibility to DNA methylation changes, CpG islands are essential targets for environmental epigenetic studies and potential biomarkers for disease susceptibility [72]. The findings in Table 5 indicate that the EFI captures the effect of environmental factors on DNA methylation.

### 3. Discussion

In our study, we proposed using the EFI to quantify the impact of environmental factors on DNA methylation from 245 pairs of Japanese monozygotic twins. The EFI calculates the difference in methylation levels between older and younger twin pairs, assuming that the more prolonged lifetime exposure to environmental factors in older twins would result in more pronounced methylation changes. This differential exposure underpins our hypothesis that the more significant the discrepancy in methylation between the age groups, the stronger the environmental influence.

Using Storey's FDR statistical method, we identified 22,752 out of 481,190 methylation sites on the Infinium HumanMethylation450 BeadChip as environmentally sensitive. The analysis of these sensitive sites, focusing on their association with diseases, cancer methylation markers, and CpG island features, further substantiated the efficacy of the EFI.

Our analysis revealed a significant correlation between environmentally sensitive methylation sites and disease development, underscoring the EFI's potential to advance our understanding of disease prevention and treatment mechanisms. Furthermore, when evaluating known methylation markers for four types of cancer, we found that 85% of these markers were among the significant sites identified by the EFI, suggesting its robustness in assessing the influence of environmental factors on methylation. The evaluation of CpG island features revealed their heightened sensitivity to environmental factors, highlighting the importance of these regions in environmental epigenetics research and their potential as biomarkers for disease susceptibility.

The challenge of distinguishing between genetic and environmental influences is particularly formidable in epigenetics, where it intersects with the broader objectives of understanding human diseases and developing therapeutic interventions. Historically, twin studies have been instrumental in dissecting the genetic and environmental contributions to phenotypic variance [54,73]. These studies have not only been applied to estimate the impact of these factors on DNA methylation at gene-specific levels, as evidenced by research from Wong [74], but also across the entire genome, as demonstrated in studies by Kuratomi [75], Kaminsky [76], and Rakyan [77]. This body of work has significantly



deepened our understanding of the regulatory processes behind epigenetic variation, unraveling the intertwined contributions of epigenetic mechanisms, environmental factors, and genetic variance to complex traits, alongside stochastic elements [53]. The insights gleaned from these studies are invaluable for grasping the intricacies of disease development and evolution [78] and for informing future epigenetic-based strategies to combat complex diseases.

Within this context of rich academic heritage, our study introduces EFI, a novel metric computed using twin subjects. The EFI embodies a significant leap forward, transcending traditional epigenetic inquiries to enhance our grasp of disease mechanisms and potential therapeutic avenues. This methodology allows us to more accurately isolate and quantify the impact of environmental factors on DNA methylation, providing crucial insights that are potentially transformative for identifying the underlying causes of human diseases and crafting new treatment strategies. Our approach thus not only enhances the precision of our epigenetic analysis but offers a powerful tool that transcends traditional research boundaries, promising substantial advancements in medical research and the development of therapeutic interventions.

The present study had some limitations. All the twin subjects were Japanese. Because of the lack of racial and regional diversity with respect to environmental factors, potential data bias may impact the numerator and denominator of the EFI. Predicting the effect of diversity on EFI values remains a challenge. Therefore, collecting data on twins of various races from different regions is essential to determine the effect of environmental factors on the EFI when using diverse populations.

In this study, we divided the twin pairs into two groups based on the median age of 53 years. This decision was made considering several factors, including the distribution of ages among our subjects and the broader context of age distribution, both globally and in Japan.

Firstly, it is crucial to note that average age and median age differ by country. According to the United Nations' World Population Prospects 2022, the global average age in 2023 is approximately 30.4 years, and the median age is about 31.0 years. For Japan, the average age is 48.4 years, and the median age is 48.6 years. Given that our study's subjects are from Japan, it was not just pertinent, but a sound and well-grounded decision to align our division with these national statistics. By selecting a median age close to Japan's average, we ensured a representative and balanced division of our sample.

The choice of the median age is crucial for statistical robustness. It ensures that the number of subjects in each group is approximately equal, which is essential for minimizing bias and enhancing the validity of comparative analyses. We conducted sensitivity analyses using different cutoff points to further validate our choice. We explored divisions at the 25th percentile (35 years) and the 75th percentile (66 years) and found that the correlation coefficients between the EFI values and these age cutoffs were 0.81 and 0.71, respectively. These high correlation values indicate that our findings are relatively sensitive to the exact cutoff point, underscoring the robustness of our methodology.

We acknowledge that differences could influence the Environmental Factor Index (EFI) in terms of cell type composition, a known confounding factor in DNA methylation analyses. Although correction for cell type distribution using leukocyte composition data, such as through the Houseman algorithm [79], is theoretically possible, the dataset used in this study had limitations. Specifically, Osaka Twin Research Group collected the data several years ago, and complete leukocyte composition information was only available for some samples. Consequently, we could not apply cell type correction across the entire dataset. Future studies would benefit from addressing this limitation by incorporating cell type composition data to more accurately assess the specific impact of environmental factors on DNA methylation. Despite this limitation, the EFI offers valuable insights into methylation variability driven by environmental influences while acknowledging the potential confounding effect of cell type distribution.

To explore potential overlaps between the Environmental Factor Index (EFI) and conventional age acceleration measures, such as Horvath’s epigenetic clock algorithms, we analyzed the EFI values associated with the DNA methylation sites used in both the Horvath2013 (353 sites) [35] and Horvath2018 (391 sites) clocks [39]. As shown in Table 6, the average EFI for the Horvath2018 sites was 1.29 with a variance of 0.33, while for the Horvath2013 sites, the average EFI was 1.26 with a variance of 0.34. These averages and variances were higher than those observed in the overall dataset, which included all 481,190 sites (average EFI = 1.18; variance = 0.28). Additionally, the proportion of statistically significant sites within the Horvath clocks was considerably higher than in the overall dataset: 338 out of 391 sites in Horvath2018 and 334 out of 353 sites in Horvath2013 were statistically significant, compared to 22,752 out of 481,190 in the entire dataset.

**Table 6.** EFI of the Horvath 2018 clock sites and the Horvath 2013 clock sites.

	Horvath 2018	Horvath 2013	All Sites
# sites	391	353	481,190
# significant sites	338	334	22,752
Mean of EFI	1.29	1.26	1.18
Std. of EFI	0.33	0.34	0.28

These results indicate that many of the DNA methylation sites used in the Horvath epigenetic clocks are also significant in the EFI analysis. This overlap is expected, given that the Horvath epigenetic clocks use a linear formula to estimate age, relying on DNA methylation sites whose values are likely to change solely in response to the environmental factor of aging. The fact that the significant sites in the EFI analysis include the sites of the Horvath epigenetic clocks suggests that the EFI is consistent with previous analytical findings. This alignment with established epigenetic age-related sites supports the validity of the EFI in capturing environmentally driven changes in DNA methylation.

One limitation of this study is the need for a more direct analysis linking the CpG sites identified by the Environmental Factor Index (EFI) to diseases or environmental exposure/phenotypes. The dataset used in this study consisted solely of DNA methylation data and did not include detailed participant attributes, such as health conditions or environmental exposure histories. As a result, assessing the direct associations between the EFI and specific diseases or environmental factors was not feasible.

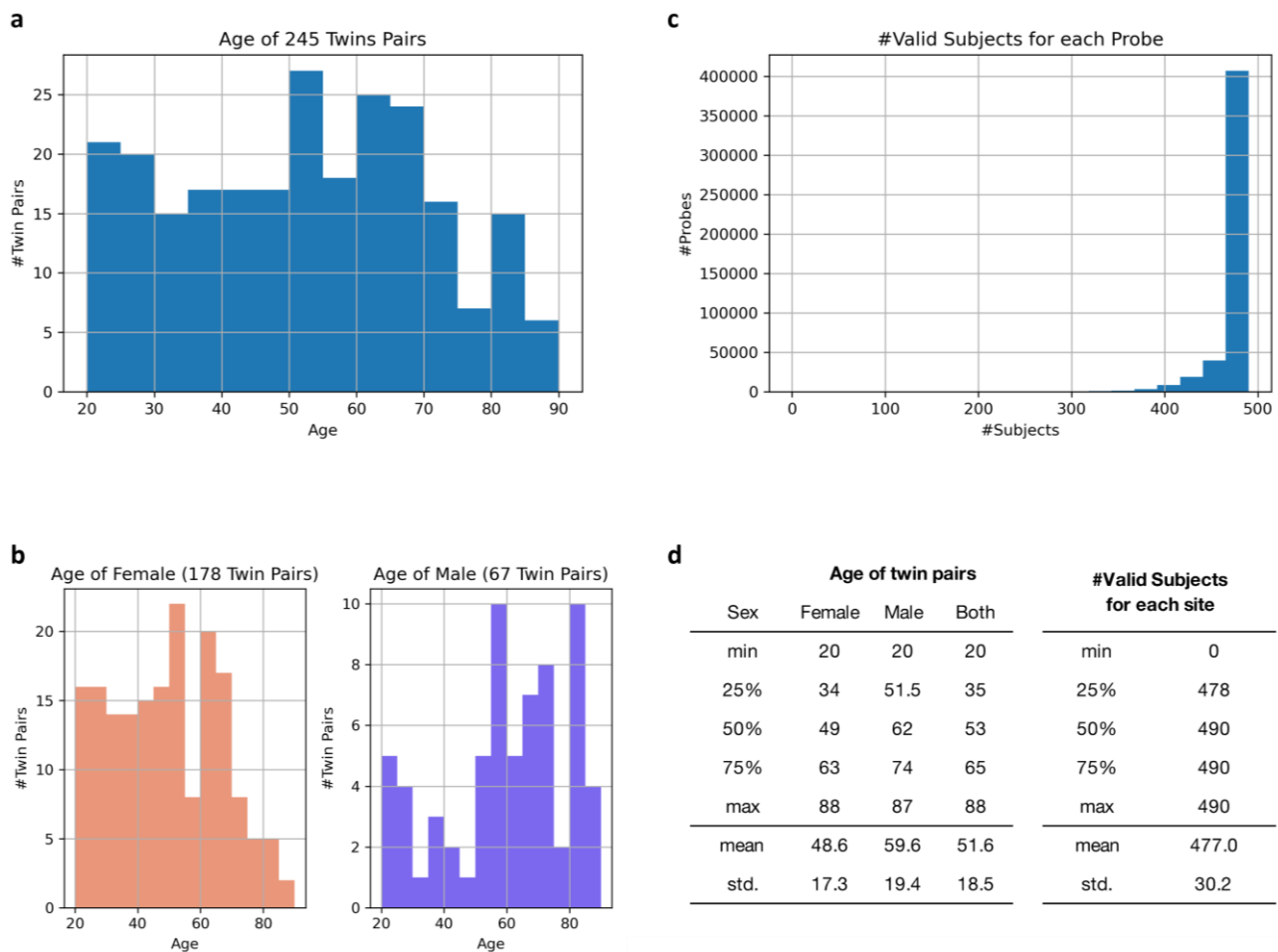
Despite this limitation, the primary objective of this study was to evaluate the variability in DNA methylation caused by environmental factors, and the EFI successfully identified methylation sites exhibiting significant changes. However, to fully understand the implications of these findings, future research should integrate DNA methylation data with detailed participant attributes. Collecting comprehensive data, including participant health conditions and environmental exposure histories, and combining these with disease-related databases will enable a more thorough exploration of the relationships between the EFI and diseases or environmental factors. Such efforts will further elucidate the role of environmental influences on epigenetic regulation and their contribution to disease pathogenesis.

## 4. Materials and Methods

### 4.1. Subjects and Ethics Statement

Since 16 January 2011, 302 healthy Japanese identical twin pairs have been recruited to measure DNA methylation. Of these, 245 were monozygotic twin pairs, and the others were dizygotic. We used the monozygotic twin pairs to ensure the genetic uniformity needed for our study. The cohort consisted of 178 pairs of females and 67 pairs of males. The subjects were selected from a registry established by the Center for Twin Research at Osaka University [80,81]. Written informed consent was obtained from all subjects before

inclusion in the study. The Ethics Committee of Osaka University approved the study protocol (No. 269). The ages of the subjects are shown in Figure 3a,b. Blood samples were collected from the subjects at 9:00 a.m. after fasting for 12 h. The subjects also underwent a clinical examination and completed a health questionnaire. The examinations were conducted on the same day for each pair of twins. The QIAamp DNA Mini Kit was used to isolate genomic DNA from the peripheral blood mononuclear cells. The zygosity of the twins was confirmed by perfectly matching 15 short tandem repeat loci using a PowerPlex® 16 System (Promega, Madison, WI, USA).



**Figure 3.** Statistics related to the subjects and data collected for the DNA methylation probes. **(a)** Age distribution of the twin pairs. The  $x$ -axis represents the age, and the  $y$ -axis represents the number of pairs. **(b)** Similar histograms showing the age distribution of the twin pairs differentiated by sex. The axes are the same as in **(a)**. **(c)** Histogram illustrating how many valid observations were collected for each DNA methylation probe. The  $x$ -axis shows the number of valid observations, and the  $y$ -axis represents the number of probes. **(d)** Representative ages of the subjects (the left table) and the number of valid subjects for each probe (the right table). The values in the right table represent basic statistics of DNA methylation probes when sorted in ascending order by the number of valid subjects.

#### 4.2. Methylation Sites

DNA methylation levels were analyzed at specific locations in the genome, known as methylation sites. The Infinium HumanMethylation450 BeadChip Kit (Illumina, San Diego, CA, USA) was used to examine 482,421 methylation sites in each sample at single-nucleotide resolution using 0.5  $\mu$ g of high-quality genomic DNA. The chip consisted of two bead types per locus at each site. The raw data obtained were analyzed using Genome Studio software (Illumina). The fluorescence intensity ratios between the two bead types

were calculated as follows: 0 indicated that the site was not methylated and 1 indicated complete methylation. A peak-based correction method [82] was used to normalize the raw data and filter out invalid reads such as null and unreliable probes. Specifically, methylation sites with fewer than 245 valid probes out of 490 subjects were excluded from further analysis, ensuring that only reliable data were retained. After this filtering process, 481,190 methylation sites remained, and their statistics are presented in Figure 3c,d.

#### 4.3. Notation

The following notations were used to measure the impact of environmental factors on each methylation site  $s$ .

$m_{i,s}$ : methylation levels of a methylation site  $s$  in a subject  $i$ . The site  $s$  was omitted if it was self-explanatory.

$T$ : a set of twin pairs included in the study. The set is represented as  $T = \{(i, j)\}$ , where  $i$  and  $j$  are the subjects of the twin pair. If  $(i, j)$  is in  $T$ , then  $(j, i)$  is not in  $T$ .

$T_s$ : a set of valid twin pairs on methylation site  $s$ .  $T_s$  is a subset of  $T$ , and only included twin pairs in which the methylation levels of site  $s$  in the subjects  $i$  and  $j$  were valid.

$D_s(T_s)$ : a set of differences in methylation levels of a methylation site  $s$  between the twins in the twin pair  $T_s$ . This is represented as  $D_s(T_s) = \{(m_{i,s} - m_{j,s}) \mid (i, j) \text{ in } T_s\}$ .

SD ( $D$ ): the standard deviation of a set of differences  $D$ .

#### 4.4. Environmental Factor Index (EFI)

To determine the strength of the influence of environmental factors on methylation sites, we compared the distribution of methylation intensities between elderly and young twins. We divided the twin pairs ( $T$ ) into two groups based on the median age of the subject, 53 years. The elderly set ( $T_{elderly}$ ) included  $\geq 53$ -year-olds and consisted of 124 twin pairs. The younger set ( $T_{young}$ ) included  $< 53$ -year-old twins and consisted of 121 pairs. Theoretically, the elderly are exposed to environmental factors for longer compared with the young pairs. Thus, the difference between the two groups indicates the degree of influence of environmental factors. We defined the EFI equation to measure this influence on the methylation site.

$$EFI_s = \frac{\text{Std}(D_s(T_{elderly}))}{\text{Std}(D_s(T_{young}))}$$

where  $T_{elderly}$  is a set of elderly twin pairs and  $T_{young}$  is a set of younger twin pairs.

The numerator of this Equation represents the distribution of methylation intensity differences in the elderly group, whereas the denominator represents the distribution in the younger group. The period affected by environmental factors is shorter in the denominator and longer in the numerator. Therefore, the EFI is a tool that allows us to deduce the impact of environmental factors on DNA methylation's variability over time. A higher EFI indicates an amplification of differences due to environmental exposure, while a lower EFI hints at a reduction in these differences.

When  $EFIs = 1$ , the impact of environmental factors on DNA methylation levels is neutral to age. This means that the variability in DNA methylation differences between twins is the same in the older and younger age groups. When  $EFIs > 1$  is greater than 1, this indicates that the variability in DNA methylation differences between twins is more significant in the older age group compared to the younger age group. This suggests that continuous exposure to environmental factors over time amplifies the differences in DNA methylation levels between twin pairs.  $EFIs < 1$  indicates that the variability in DNA methylation differences between twins is smaller in the older age group compared to the younger age group. This suggests that continuous exposure to environmental factors over time decreases the differences in DNA methylation levels between twin pairs.

#### 4.5. Statistical Analysis of EFI

We aimed to investigate how environmental factors influence DNA methylation in monozygotic twins. To achieve this, we developed the Environmental Factor Index (EFI),

which quantifies the variability in DNA methylation levels between younger and older twins, thus isolating the impact of environmental exposure over time. By comparing the differences in methylation intensities between these two groups, the EFI enables us to assess how environmental factors contribute to methylation changes at specific genomic sites.

Figure 2 presents statistics on EFI. Figure 2a shows a scatter chart of the standard deviations (SD) for young and elderly twins. Each dot on the chart represents a DNA methylation site. The line indicates that the values of the two axes are equivalent. Interestingly, 76.5% of the sites are above the line, indicating that the differences in the degree of methylation between the twins increased, specifically reflecting that the standard deviation of methylation in the elderly group is more significant than that in the younger group. The sites below the line indicate a decreasing difference between twins in a time-dependent manner, specifically reflecting that the standard deviation in the younger group is more significant than that in the elderly group.

We conducted a test of the EFI, where the denominator is the SD of young twin pairs and the numerator is the SD of elderly twin pairs. Therefore, to test the EFI, we utilized Levene's test, which assesses the equality of variances between the two groups, to examine whether there is a significant difference between the standard deviations of these two populations. Specifically, we formulated a null hypothesis that the variances in the denominator and numerator are equal. We performed these calculations using Python 3, specifically employing the SciPy library for Levene's test. Given that 481,190 DNA methylation sites are subject to testing, it is necessary to address the issue of multiple comparisons. Therefore, we applied Storey's FDR approach [83]. To determine the significance of the methylation sites, we used a significance level of 1%.

Figure 2b shows the results of Levene's test for each methylation site using the multiple testing correction. The line represents the probability of the test after the correction of Storey's FDR ( $q$ -value). The dotted horizontal line indicates the significance level of 1%. The number of sites in which the null hypothesis was rejected was 22,752 and the ratio to the total number of observed methylation sites (481,190 sites) was 4.7%. For convenience, we refer to the methylation sites in which the null hypotheses were rejected by Storey's FDR as significant sites.

To validate the EFI statistically, we performed Levene's test to compare the variability in DNA methylation levels between younger and older twins. This analysis determines whether the observed differences in methylation variability are statistically significant, thus confirming the influence of environmental factors on DNA methylation over time.

Figure 2c shows a scatter chart of the EFI and Storey's FDR results, analyzed using a Levene's test, for young and elderly twins. Each dot represents a methylation site, with logarithmic scales used for the  $x$ - and  $y$ -axes. A depression is evident in the chart at  $EFI = 1$ . Because the denominator and numerator of the EFI are identical at this value, the  $p$ -value for Levene's test is necessarily close to 1. In this scatter chart, methylation sites with a small FDR are more prevalent on the right side of this Figure, where  $EFI > 1$ . Of the 22,752 methylation sites identified as significant, 22,568 had an  $EFI > 1$ , whereas only 94 sites had an  $EFI < 1$ . For our interpretation, we designated the significant sites with  $EFI$  values  $> 1$  as "elderly-significant sites" and those with  $< 1$  as "young-significant sites".

The histogram in Figure 2d shows the EFI distribution with logarithmic  $x$ - and  $y$ -axes. Of the 481,190 methylation sites, 112,868 (23%) had an  $EFI < 1$ , and 368,322 (77%) sites had an  $EFI > 1$ . The median  $EFI$  value was 1.137, indicating that the difference in the degree of DNA methylation between twins may increase over time.

#### 4.6. Statistical Analysis of Disease Association

We identified the methylation sites and associated genes selected by the EFI and investigated their relationship with disorders using the MalaCards database. Following this, we performed a coincidence test to confirm that the associations discovered by the EFI were not due to random chance. The null hypothesis stated that the elite genes appeared by chance.



MalaCards is a comprehensive database that compiles information on human diseases, including genetic associations, pathways, and relevant research. Some of the genes listed in MalaCards are classified as “elite genes”. These elite genes, which we found to be of the utmost importance, are genes that have strong evidence linking them to specific diseases based on high-confidence sources such as curated databases, clinical studies, or genome-wide association studies (GWAS). The “elite” designation indicates that these genes are strongly associated with certain disorders, making them important candidates for further investigation in disease-related studies.

The MalaCards database does not publish the number or ratio of elite genes, which is necessary information for the coincidence test. Therefore, we independently gathered the relevant data by randomly sampling 100 sites from an array of 481,190 sites on BeadChip. To assess the statistical significance of this association, we applied a two-tailed Fisher’s exact test with a significance level of 1%.

## 5. Conclusions

Our study has pioneered use of the EFI, marking a significant stride in the epigenetic domain by quantifying the environmental impacts on DNA methylation with unparalleled precision. This novel metric utilizes an extensive dataset derived from 245 monozygotic twin pairs. It examines 481,190 methylation sites, establishing a robust framework that enhances our understanding of the intricate relationship between the environment and the epigenome.

The identification of 22,752 DNA methylation sites significantly influenced by environmental factors using the Environmental Factor Index (EFI) highlights the tool’s sensitivity and capacity to elucidate the complex dynamics of epigenetic modifications. This accomplishment is particularly significant given the long-standing challenge of detecting these subtle influences within the complex epigenome. Moreover, the analysis conducted with the EFI reveals that although individual environmentally sensitive methylation sites might only slightly impact phenotypic traits, their aggregated effect within genes significantly correlates with phenotypic expression. This insight enriches our understanding of the cumulative influence of the environment on gene regulation and expression.

Moreover, by applying the EFI to the MalaCards database, our study illuminated the potential connections between environmentally influenced methylation sites and various disorders, notably cancer, suggesting the EFI’s utility in identifying epigenetic markers of disease. This application validates the EFI’s relevance and opens new research avenues for exploring the epigenetic underpinnings of various diseases and conditions.

The introduction of the EFI represents a transformative development in epigenetic research, offering a new lens through which we can explore the complex interplay between the genome, the epigenome, and environmental factors. As we move forward, the EFI stands to significantly enhance our understanding of epigenetic regulation, shed light on the mechanisms underlying disease, and inform the development of novel diagnostic and therapeutic approaches. Our findings affirm the potential of the EFI as a cornerstone in future epigenetic investigations, heralding a new era of precision in quantifying environmental influences on the epigenome.

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/epigenomes8040044/s1>, Marker.pdf, EFI.xlsx, EFGene.csv. Marker.pdf: The file includes DNA methylation marker genes for four cancers and the number of EFI significant sites. EFI.xlsx: The file includes the EFI and Storey’s FDR for each of the 481,190 methylation sites. Each methylation site has a probe ID of the HumanMethylation450 BeadChip (GPL 18809 @ Gene Expression Omnibus) and a gene symbol. The csv file has four columns, as follows: 1. Probe ID; 2. Gene symbol; 3. EFI; 4. FDR. EFGene.csv: The file includes environmentally sensitive gene symbols. The file has three columns, as follows: 1. Gene symbol; 2. Number of elderly significant probes; 3. Number of young significant probes.



**Author Contributions:** Data curation, Y.T., Osaka Twin Research Group and M.W.; funding acquisition, Osaka Twin Research Group and M.W.; methodology, Y.T.; project administration, Osaka Twin Research Group and M.W.; resources, Osaka Twin Research Group and M.W.; software, Y.T.; writing—original draft, Y.T.; writing—review and editing, Y.T. All authors have read and agreed to the published version of the manuscript.

**Funding:** This study was supported by University Grants from the Japanese Ministry of Education, Culture, Sports, Science, and Technology and the Kansai University Fund for Domestic and Overseas Research Fund, 2020 and 2021.

**Institutional Review Board Statement:** The study was conducted in accordance with the Declaration of Helsinki, and approved by the Ethics Committee of Osaka University (No. 269, 24 August 2010).

**Informed Consent Statement:** Informed consent was obtained from all subjects involved in the study. Written informed consent has been obtained from the patients to publish this paper.

**Data Availability Statement:** The methylation levels of each site for each subject that support the results of this study are not publicly available. The small number of twin participants, coupled with the unique identifying nature of twin status, facilitates easier identification of individuals. Therefore, due to the Center for Twin Research's policies to protect the individual's privacy, the data have not been made publicly accessible. However, upon reasonable request, data may be obtained with the permission of the Center for Twin Research, Osaka University Graduate School of Medicine (research@twin.med.osaka-u.ac.jp).

**Conflicts of Interest:** The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

## References

- Berger, S.L.; Kouzarides, T.; Shiekhattar, R.; Shilatifard, A. An operational definition of epigenetics. *Genes Dev.* **2009**, *23*, 781–783. [[CrossRef](#)]
- Dunn, D.; Smith, J. The occurrence of 6-methylaminopurine in deoxyribonucleic acids. *Biochem. J.* **1958**, *68*, 627. [[CrossRef](#)]
- Vanyushin, B.; Tkacheva, S.; Belozersky, A. Rare bases in animal DNA. *Nature* **1970**, *225*, 948–949. [[CrossRef](#)]
- Ehrlich, M.; Gama-Sosa, M.A.; Carreira, L.H.; Ljungdahl, L.G.; Kuo, K.C.; Gehrke, C.W. DNA methylation in thermophilic bacteria: N 4-methylcytosine, 5-methylcytosine, and N 5 methyladenine. *Nucleic Acids Res.* **1985**, *13*, 1399–1412. [[CrossRef](#)]
- Ratel, D.; Ravanat, J.L.; Berger, F.; Wion, D. N6-methyladenine: The other methylated base of DNA. *Bioessays* **2006**, *28*, 309–315. [[CrossRef](#)]
- Wu, T.P.; Wang, T.; Seetin, M.G.; Lai, Y.; Zhu, S.; Lin, K.; Liu, Y.; Byrum, S.D.; Mackintosh, S.G.; Zhong, M. DNA methylation on N 6-adenine in mammalian embryonic stem cells. *Nature* **2016**, *532*, 329–333. [[CrossRef](#)]
- Goll, M.G.; Bestor, T.H. Eukaryotic cytosine methyltransferases. *Annu. Rev. Biochem.* **2005**, *74*, 481–514. [[CrossRef](#)]
- Bostick, M.; Kim, J.K.; Estève, P.-O.; Clark, A.; Pradhan, S.; Jacobsen, S.E. UHRF1 plays a role in maintaining DNA methylation in mammalian cells. *Science* **2007**, *317*, 1760–1764. [[CrossRef](#)]
- Jones, P.A. Functions of DNA methylation: Islands, start sites, gene bodies and beyond. *Nat. Rev. Genet.* **2012**, *13*, 484–492. [[CrossRef](#)]
- Smith, Z.D.; Meissner, A. DNA methylation: Roles in mammalian development. *Nat. Rev. Genet.* **2013**, *14*, 204–220. [[CrossRef](#)]
- Yin, Y.; Morgunova, E.; Jolma, A.; Kaasinen, E.; Sahu, B.; Khund-Sayeed, S.; Das, P.K.; Kivioja, T.; Dave, K.; Zhong, F. Impact of cytosine methylation on DNA binding specificities of human transcription factors. *Science* **2017**, *356*, eaaj2239. [[CrossRef](#)]
- Bender, C.M.; Gonzalgo, M.L.; Gonzales, F.A.; Nguyen, C.T.; Robertson, K.D.; Jones, P.A. Roles of cell division and gene transcription in the methylation of CpG islands. *Mol. Cell. Biol.* **1999**, *19*, 6690–6698. [[CrossRef](#)]
- Maunakea, A.K.; Chepelev, I.; Cui, K.; Zhao, K. Intragenic DNA methylation modulates alternative splicing by recruiting MeCP2 to promote exon recognition. *Cell Res.* **2013**, *23*, 1256–1269. [[CrossRef](#)]
- Maunakea, A.K.; Nagarajan, R.P.; Bilenky, M.; Ballinger, T.J.; D'Souza, C.; Fouse, S.D.; Johnson, B.E.; Hong, C.; Nielsen, C.; Zhao, Y. Conserved role of intragenic DNA methylation in regulating alternative promoters. *Nature* **2010**, *466*, 253–257. [[CrossRef](#)]
- Robertson, K.D.; Wolffe, A.P. DNA methylation in health and disease. *Nat. Rev. Genet.* **2000**, *1*, 11–19. [[CrossRef](#)]
- Müller, H.M.; Fiegl, H.; Widschwendter, A.; Widschwendter, M. Prognostic DNA methylation marker in serum of cancer patients. *Ann. N. Y. Acad. Sci.* **2004**, *1022*, 44–49. [[CrossRef](#)]
- Sigin, V.O.; Kalinkin, A.I.; Kuznetsova, E.B.; Simonova, O.A.; Chesnokova, G.G.; Litviakov, N.V.; Slonimskaya, E.M.; Tsyganov, M.M.; Ibragimova, M.K.; Volodin, I.V. DNA methylation markers panel can improve prediction of response to neoadjuvant chemotherapy in luminal B breast cancer. *Sci. Rep.* **2020**, *10*, 9239. [[CrossRef](#)]
- Xu, W.; Xu, M.; Wang, L.; Zhou, W.; Xiang, R.; Shi, Y.; Zhang, Y.; Piao, Y. Integrative analysis of DNA methylation and gene expression identified cervical cancer-specific diagnostic biomarkers. *Signal Transduct. Target. Ther.* **2019**, *4*, 55. [[CrossRef](#)]

19. Lofton-Day, C.; Model, F.; DeVos, T.; Tetzner, R.; Distler, J.; Schuster, M.; Song, X.; Lesche, R.; Liebenberg, V.; Ebert, M. DNA methylation biomarkers for blood-based colorectal cancer screening. *Clin. Chem.* **2008**, *54*, 414–423. [[CrossRef](#)]
20. Tänzer, M.; Balluff, B.; Distler, J.; Hale, K.; Leodolter, A.; Röcken, C.; Molnar, B.; Schmid, R.; Lofton-Day, C.; Schuster, T. Performance of epigenetic markers SEPT9 and ALX4 in plasma for detection of colorectal precancerous lesions. *PLoS ONE* **2010**, *5*, e9061. [[CrossRef](#)]
21. Kim, M.S.; Lee, J.; Sidransky, D. DNA methylation markers in colorectal cancer. *Cancer Metastasis Rev.* **2010**, *29*, 181–206. [[CrossRef](#)]
22. Diaz-Lagares, A.; Mendez-Gonzalez, J.; Hervas, D.; Saigi, M.; Pajares, M.J.; Garcia, D.; Crujeiras, A.B.; Pio, R.; Montuenga, L.M.; Zulueta, J. A novel epigenetic signature for early diagnosis in lung cancer. *Clin. Cancer Res.* **2016**, *22*, 3361–3371. [[CrossRef](#)]
23. Yan, P.; Yang, X.; Wang, J.; Wang, S.; Ren, H. A novel CpG island methylation panel predicts survival in lung adenocarcinomas. *Oncol. Lett.* **2019**, *18*, 1011–1022. [[CrossRef](#)]
24. Li, M.; Zhang, C.; Zhou, L.; Li, S.; Cao, Y.J.; Wang, L.; Xiang, R.; Shi, Y.; Piao, Y. Identification and validation of novel DNA methylation markers for early diagnosis of lung adenocarcinoma. *Mol. Oncol.* **2020**, *14*, 2744–2758. [[CrossRef](#)]
25. Rakyan, V.K.; Beyan, H.; Down, T.A.; Hawa, M.I.; Maslau, S.; Aden, D.; Daunay, A.; Busato, F.; Mein, C.A.; Manfras, B. Identification of type 1 diabetes-associated DNA methylation variable positions that precede disease diagnosis. *PLoS Genet.* **2011**, *7*, e1002300. [[CrossRef](#)]
26. Davegårdh, C.; García-Calzón, S.; Bacos, K.; Ling, C. DNA methylation in the pathogenesis of type 2 diabetes in humans. *Mol. Metab.* **2018**, *14*, 12–25. [[CrossRef](#)]
27. Ahmed, S.A.H.; Ansari, S.A.; Mensah-Brown, E.P.; Emerald, B.S. The role of DNA methylation in the pathogenesis of type 2 diabetes mellitus. *Clin. Epigenetics* **2020**, *12*, 104. [[CrossRef](#)]
28. Bechtel, W.; McGoohan, S.; Zeisberg, E.M.; Müller, G.A.; Kalbacher, H.; Salant, D.J.; Müller, C.A.; Kalluri, R.; Zeisberg, M. Methylation determines fibroblast activation and fibrogenesis in the kidney. *Nat. Med.* **2010**, *16*, 544–550. [[CrossRef](#)]
29. Chu, A.Y.; Tin, A.; Schlosser, P.; Ko, Y.-A.; Qiu, C.; Yao, C.; Joehanes, R.; Grams, M.E.; Liang, L.; Gluck, C.A. Epigenome-wide association studies identify DNA methylation associated with kidney function. *Nat. Commun.* **2017**, *8*, 1286. [[CrossRef](#)]
30. Coit, P.; Jeffries, M.; Altorok, N.; Dozmorov, M.G.; Koelsch, K.A.; Wren, J.D.; Merrill, J.T.; McCune, W.J.; Sawalha, A.H. Genome-wide DNA methylation study suggests epigenetic accessibility and transcriptional poisoning of interferon-regulated genes in naive CD4+ T cells from lupus patients. *J. Autoimmun.* **2013**, *43*, 78–84. [[CrossRef](#)]
31. Imgenberg-Kreuz, J.; Almlöf, J.C.; Leonard, D.; Alexsson, A.; Nordmark, G.; Eloranta, M.-L.; Rantapää-Dahlqvist, S.; Bengtsson, A.A.; Jönsen, A.; Padyukov, L. DNA methylation mapping identifies gene regulatory effects in patients with systemic lupus erythematosus. *Ann. Rheum. Dis.* **2018**, *77*, 736–743. [[CrossRef](#)] [[PubMed](#)]
32. Imgenberg-Kreuz, J.; Almlöf, J.C.; Leonard, D.; Sjöwall, C.; Syvänen, A.-C.; Rönnblom, L.; Sandling, J.K.; Nordmark, G. Shared and Unique Patterns of DNA Methylation in Systemic Lupus Erythematosus and Primary Sjögren’s Syndrome. *Front. Immunol.* **2019**, *10*, 1686. [[CrossRef](#)] [[PubMed](#)]
33. Edris, A.; den Dekker, H.T.; Melén, E.; Lahousse, L. Epigenome-wide association studies in asthma: A systematic review. *Clin. Exp. Allergy* **2019**, *49*, 953–968. [[CrossRef](#)] [[PubMed](#)]
34. Alag, A. Machine learning approach yields epigenetic biomarkers of food allergy: A novel 13-gene signature to diagnose clinical reactivity. *PLoS ONE* **2019**, *14*, e0218253. [[CrossRef](#)]
35. Horvath, S. DNA methylation age of human tissues and cell types. *Genome Biol.* **2013**, *14*, 3156. [[CrossRef](#)]
36. Hannum, G.; Guinney, J.; Zhao, L.; Zhang, L.; Hughes, G.; Sadda, S.; Klotzle, B.; Bibikova, M.; Fan, J.-B.; Gao, Y. Genome-wide methylation profiles reveal quantitative views of human aging rates. *Mol. Cell* **2013**, *49*, 359–367. [[CrossRef](#)]
37. Weidner, C.I.; Lin, Q.; Koch, C.M.; Eisele, L.; Beier, F.; Ziegler, P.; Bauerschlag, D.O.; Jöckel, K.-H.; Erbel, R.; Mühleisen, T.W. Aging of blood can be tracked by DNA methylation changes at just three CpG sites. *Genome Biol.* **2014**, *15*, R24. [[CrossRef](#)]
38. Levine, M.E.; Lu, A.T.; Quach, A.; Chen, B.H.; Assimes, T.L.; Bandinelli, S.; Hou, L.; Baccarelli, A.A.; Stewart, J.D.; Li, Y. An epigenetic biomarker of aging for lifespan and healthspan. *Aging* **2018**, *10*, 573. [[CrossRef](#)]
39. Horvath, S.; Raj, K. DNA methylation-based biomarkers and the epigenetic clock theory of ageing. *Nat. Rev. Genet.* **2018**, *19*, 371–384. [[CrossRef](#)]
40. Kim, M.K.; Sasaki, S.; Otani, T.; Tsugane, S.; Japan Public Health Center-based Prospective Study Group. Dietary patterns and subsequent colorectal cancer risk by subsite: A prospective cohort study. *Int. J. Cancer* **2005**, *115*, 790–798. [[CrossRef](#)]
41. Wang, K.; Gaitsch, H.; Poon, H.; Cox, N.J.; Rzhetsky, A. Classification of common human diseases derived from shared genetic and environmental determinants. *Nat. Genet.* **2017**, *49*, 1319. [[CrossRef](#)] [[PubMed](#)]
42. Lichtenstein, P.; Holm, N.V.; Verkasalo, P.K.; Iliadou, A.; Kaprio, J.; Koskenvuo, M.; Pukkala, E.; Skytthe, A.; Hemminki, K. Environmental and heritable factors in the causation of cancer—Analyses of cohorts of twins from Sweden, Denmark, and Finland. *N. Engl. J. Med.* **2000**, *343*, 78–85. [[CrossRef](#)] [[PubMed](#)]
43. Keil, K.P.; Lein, P.J. DNA methylation: A mechanism linking environmental chemical exposures to risk of autism spectrum disorders? *Environ. Epigenetics* **2016**, *2*, dvv012. [[CrossRef](#)] [[PubMed](#)]
44. Seong, K.-H.; Maekawa, T.; Ishii, S. Inheritance of Stress-Induced Epigenetic Changes Mediated by the ATF-2 Family of Transcription Factors. In *Stress-Induced Mutagenesis*; Springer: Berlin/Heidelberg, Germany, 2013; pp. 103–118.
45. Radford, E.J.; Ito, M.; Shi, H.; Corish, J.A.; Yamazawa, K.; Isganaitis, E.; Seisenberger, S.; Hore, T.A.; Reik, W.; Erkek, S. In utero undernourishment perturbs the adult sperm methylome and is linked to metabolic disease transmission. *Science* **2014**, *345*, 1255903. [[CrossRef](#)]

46. Bell, J.T.; Spector, T.D. A twin approach to unraveling epigenetics. *Trends Genet.* **2011**, *27*, 116–125. [[CrossRef](#)]
47. Egger, G.; Liang, G.; Aparicio, A.; Jones, P.A. Epigenetics in human disease and prospects for epigenetic therapy. *Nature* **2004**, *429*, 457–463. [[CrossRef](#)]
48. Fujii, R.; Sato, S.; Tsuboi, Y.; Cardenas, A.; Suzuki, K. DNA methylation as a mediator of associations between the environment and chronic diseases: A scoping review on application of mediation analysis. *Epigenetics* **2022**, *17*, 759–785. [[CrossRef](#)]
49. Burgio, E.; Piscitelli, P.; Colao, A. Environmental carcinogenesis and transgenerational transmission of carcinogenic risk: From genetics to epigenetics. *Int. J. Environ. Res. Public Health* **2018**, *15*, 1791. [[CrossRef](#)]
50. Fallet, M.; Blanc, M.; Di Criscio, M.; Antczak, P.; Engwall, M.; Bosagna, C.G.; Rüegg, J.; Keiter, S.H. Present and future challenges for the investigation of transgenerational epigenetic inheritance. *Environ. Int.* **2023**, *172*, 107776. [[CrossRef](#)]
51. Boomsma, D.; Busjahn, A.; Peltonen, L. Classical twin studies and beyond. *Nat. Rev. Genet.* **2002**, *3*, 872–882. [[CrossRef](#)]
52. Silventoinen, K.; Rokholm, B.; Kaprio, J.; Sørensen, T.I. The genetic and environmental influences on childhood obesity: A systematic review of twin and adoption studies. *Int. J. Obes.* **2010**, *34*, 29–40. [[CrossRef](#)]
53. Bell, J.T.; Saffery, R. The value of twins in epigenetic epidemiology. *Int. J. Epidemiol.* **2012**, *41*, 140–150. [[CrossRef](#)] [[PubMed](#)]
54. Gordon, L.; Joo, J.E.; Powell, J.E.; Ollikainen, M.; Novakovic, B.; Li, X.; Andronikos, R.; Cruickshank, M.N.; Conneely, K.N.; Smith, A.K. Neonatal DNA methylation profile in human twins is specified by a complex interplay between intrauterine environmental and genetic factors, subject to tissue-specific influence. *Genome Res.* **2012**, *22*, 1395–1406. [[CrossRef](#)] [[PubMed](#)]
55. Rappaport, N.; Twik, M.; Plaschkes, I.; Nudel, R.; Iny Stein, T.; Levitt, J.; Gershoni, M.; Morrey, C.P.; Safran, M.; Lancet, D. MalaCards: An amalgamated human disease compendium with diverse clinical and genetic annotation and structured search. *Nucleic Acids Res.* **2017**, *45*, D877–D887. [[CrossRef](#)] [[PubMed](#)]
56. Behnia, F.; Parets, S.E.; Kechichian, T.; Yin, H.; Dutta, E.H.; Saade, G.R.; Smith, A.K.; Menon, R. Fetal DNA methylation of autism spectrum disorders candidate genes: Association with spontaneous preterm birth. *Am. J. Obstet. Gynecol.* **2015**, *212*, 533.e1–533.e9. [[CrossRef](#)]
57. Mouat, J.S.; Li, X.; Neier, K.; Zhu, Y.; Mordaunt, C.E.; La Merrill, M.A.; Lehmler, H.J.; Jones, M.P.; Lein, P.J.; Schmidt, R.J.; et al. Networks of placental DNA methylation correlate with maternal serum PCB concentrations and child neurodevelopment. *Environ. Res.* **2023**, *220*, 115227. [[CrossRef](#)]
58. Xu, X.; Yuan, X.; Ni, J.; Guo, J.; Gao, Y.; Yin, W.; Li, F.; Wei, L.; Zhang, J. MAGI2-AS3 inhibits breast cancer by downregulating DNA methylation of MAGI2. *J. Cell. Physiol.* **2021**, *236*, 1116–1130. [[CrossRef](#)]
59. Qu, Y.; Gao, N.; Wu, T. Expression and clinical significance of SYNE1 and MAGI2 gene promoter methylation in gastric cancer. *Medicine* **2021**, *100*, e23788. [[CrossRef](#)]
60. Chang, C.C.; Wang, H.C.; Liao, Y.P.; Chen, Y.C.; Weng, Y.C.; Yu, M.H.; Lai, H.C. The feasibility of detecting endometrial and ovarian cancer using DNA methylation biomarkers in cervical scrapings. *J. Gynecol. Oncol.* **2018**, *29*, e17. [[CrossRef](#)]
61. Jones, P.A.; Issa, J.-P.J.; Baylin, S. Targeting the cancer epigenome for therapy. *Nat. Rev. Genet.* **2016**, *17*, 630. [[CrossRef](#)]
62. Dawson, M.A. The cancer epigenome: Concepts, challenges, and therapeutic opportunities. *Science* **2017**, *355*, 1147–1152. [[CrossRef](#)] [[PubMed](#)]
63. Raut, J.R.; Guan, Z.; Schrotz-King, P.; Brenner, H. Fecal DNA methylation markers for detecting stages of colorectal cancer and its precursors: A systematic review. *Clin. Epigenetics* **2020**, *12*, 122. [[CrossRef](#)] [[PubMed](#)]
64. Saxonov, S.; Berg, P.; Brutlag, D.L. A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 1412–1417. [[CrossRef](#)] [[PubMed](#)]
65. Goldberg, A.D.; Allis, C.D.; Bernstein, E. Epigenetics: A landscape takes shape. *Cell* **2007**, *128*, 635–638. [[CrossRef](#)] [[PubMed](#)]
66. Maor, G.L.; Yearim, A.; Ast, G. The alternative role of DNA methylation in splicing regulation. *Trends Genet.* **2015**, *31*, 274–280. [[CrossRef](#)]
67. Luo, C.; Hajkova, P.; Ecker, J.R. Dynamic DNA methylation: In the right place at the right time. *Science* **2018**, *361*, 1336–1340. [[CrossRef](#)]
68. Bird, A. DNA methylation patterns and epigenetic memory. *Genes Dev.* **2002**, *16*, 6–21. [[CrossRef](#)]
69. Fatemi, M.; Pao, M.M.; Jeong, S.; Gal-Yam, E.N.; Egger, G.; Weisenberger, D.J.; Jones, P.A. Footprinting of mammalian promoters: Use of a CpG DNA methyltransferase revealing nucleosome positions at a single molecule level. *Nucleic Acids Res.* **2005**, *33*, e176. [[CrossRef](#)]
70. Irizarry, R.A.; Ladd-Acosta, C.; Wen, B.; Wu, Z.; Montano, C.; Onyango, P.; Cui, H.; Gabo, K.; Rongione, M.; Webster, M. Genome-wide methylation analysis of human colon cancer reveals similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. *Nat. Genet.* **2009**, *41*, 178. [[CrossRef](#)]
71. Doi, A.; Park, I.-H.; Wen, B.; Murakami, P.; Aryee, M.J.; Irizarry, R.; Herb, B.; Ladd-Acosta, C.; Rho, J.; Loewer, S. Differential methylation of tissue- and cancer-specific CpG island shores distinguishes human induced pluripotent stem cells, embryonic stem cells and fibroblasts. *Nat. Genet.* **2009**, *41*, 1350–1353. [[CrossRef](#)]
72. Cortessis, V.K.; Thomas, D.C.; Levine, A.J.; Breton, C.V.; Mack, T.M.; Siegmund, K.D.; Haile, R.W.; Laird, P.W. Environmental epigenetics: Prospects for studying epigenetic mediation of exposure–response relationships. *Hum. Genet.* **2012**, *131*, 1565–1589. [[CrossRef](#)] [[PubMed](#)]
73. Van Dongen, J.; Gordon, S.D.; McRae, A.F.; Odintsova, V.V.; Mbarek, H.; Breeze, C.E.; Sugden, K.; Lundgren, S.; Castillo-Fernandez, J.E.; Hannon, E. Identical twins carry a persistent epigenetic signature of early genome programming. *Nat. Commun.* **2021**, *12*, 5618. [[CrossRef](#)]

74. Wong, C.C.Y.; Caspi, A.; Williams, B.; Craig, I.W.; Houts, R.; Ambler, A.; Moffitt, T.E.; Mill, J. A longitudinal study of epigenetic variation in twins. *Epigenetics* **2010**, *5*, 516–526. [[CrossRef](#)] [[PubMed](#)]
75. Kuratomi, G.; Iwamoto, K.; Bundo, M.; Kusumi, I.; Kato, N.; Iwata, N.; Ozaki, N.; Kato, T. Aberrant DNA methylation associated with bipolar disorder identified from discordant monozygotic twins. *Mol. Psychiatry* **2008**, *13*, 429–441. [[CrossRef](#)] [[PubMed](#)]
76. Kaminsky, Z.A.; Tang, T.; Wang, S.-C.; Ptak, C.; Oh, G.H.; Wong, A.H.; Feldcamp, L.A.; Virtanen, C.; Halfvarson, J.; Tysk, C. DNA methylation profiles in monozygotic and dizygotic twins. *Nat. Genet.* **2009**, *41*, 240–245. [[CrossRef](#)] [[PubMed](#)]
77. Rakyan, V.K.; Down, T.A.; Balding, D.J.; Beck, S. Epigenome-wide association studies for common human diseases. *Nat. Rev. Genet.* **2011**, *12*, 529–541. [[CrossRef](#)] [[PubMed](#)]
78. Feinberg, A.P.; Irizarry, R.A. Evolution in health and medicine Sackler colloquium: Stochastic epigenetic variation as a driving force of development, evolutionary adaptation, and disease. *Proc. Natl. Acad. Sci. USA* **2009**, *107*, 1757–1764. [[CrossRef](#)] [[PubMed](#)]
79. Houseman, E.A.; Accomando, W.P.; Koestler, D.C.; Christensen, B.C.; Marsit, C.J.; Nelson, H.H.; Wiencke, J.K.; Kelsey, K.T. DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinform.* **2012**, *13*, 86. [[CrossRef](#)]
80. Honda, C.; Watanabe, M.; Tomizawa, R.; Sakai, N.; Group, O.T.R. Update on Osaka University Twin Registry: An Overview of Multidisciplinary Research Resources and Biobank at Osaka University Center for Twin Research. *Twin Res. Hum. Genet.* **2019**, *22*, 597–601. [[CrossRef](#)]
81. Watanabe, M.; Honda, C.; Iwatani, Y.; Yorifuji, S.; Iso, H.; Kamide, K.; Hatazawa, J.; Kihara, S.; Sakai, N.; Watanabe, H. Within-pair differences of DNA methylation levels between monozygotic twins are different between male and female pairs. *BMC Med. Genom.* **2016**, *9*, 55. [[CrossRef](#)]
82. Dedeurwaerder, S.; Defrance, M.; Calonne, E.; Denis, H.; Sotiriou, C.; Fuks, F. Evaluation of the Infinium Methylation 450K technology. *Epigenomics* **2011**, *3*, 771–784. [[CrossRef](#)] [[PubMed](#)]
83. Storey, J.D.; Tibshirani, R. Statistical significance for genomewide studies. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 9440–9445. [[CrossRef](#)] [[PubMed](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.