

Article

# Development of Tissue-Specific Age Predictors Using DNA Methylation Data

Heeyeon Choi, Soobok Joe and Hojung Nam \* 

School of Electrical Engineering and Computer Science, Gwangju Institute of Science of Technology, Gwangju 61005, Korea; heeyeon456@naver.com (H.C.); soobok@gist.ac.kr (S.J.)

\* Correspondence: hjnam@gist.ac.kr

Received: 23 September 2019; Accepted: 1 November 2019; Published: 4 November 2019



**Abstract:** DNA methylation patterns have been shown to change throughout the normal aging process. Several studies have found epigenetic aging markers using age predictors, but these studies only focused on blood-specific or tissue-common methylation patterns. Here, we constructed nine tissue-specific age prediction models using methylation array data from normal samples. The constructed models predict the chronological age with good performance (mean absolute error of 5.11 years on average) and show better performance in the independent test than previous multi-tissue age predictors. We also compared tissue-common and tissue-specific aging markers and found that they had different characteristics. Firstly, the tissue-common group tended to contain more positive aging markers with methylation values that increased during the aging process, whereas the tissue-specific group tended to contain more negative aging markers. Secondly, many of the tissue-common markers were located in Cytosine-phosphate-Guanine (CpG) island regions, whereas the tissue-specific markers were located in CpG shore regions. Lastly, the tissue-common CpG markers tended to be located in more evolutionarily conserved regions. In conclusion, our prediction models identified CpG markers that capture both tissue-common and tissue-specific characteristics during the aging process.

**Keywords:** epigenetics; DNA methylation; age prediction; tissue-specific methylation

## 1. Introduction

Aging is often defined as an overall functional decline over time that affects all living organisms [1]. In addition, the human aging process is an important health factor that is still not fully explainable, thus understanding what happens when people age is of great interest [2,3]. In recent years, many cellular and molecular hallmarks of aging have been discovered, including cellular senescence [4], gene expression changes [5,6], and telomere attrition [7–9]. Moreover, biological markers of aging involving epigenetic changes have been studied, and DNA methylation has emerged as a promising biomarker of healthy human aging [10].

DNA methylation is a representative epigenetic modification that involves the addition of methyl groups to the DNA molecule. DNA methylation is considered to be a crucial epigenetic change because it can alter the activity of genes and is a biomarker that has been implicated in various human diseases, including cancer. Furthermore, many studies revealed that DNA methylation patterns in specific regions have been shown to change along with the aging process [11,12].

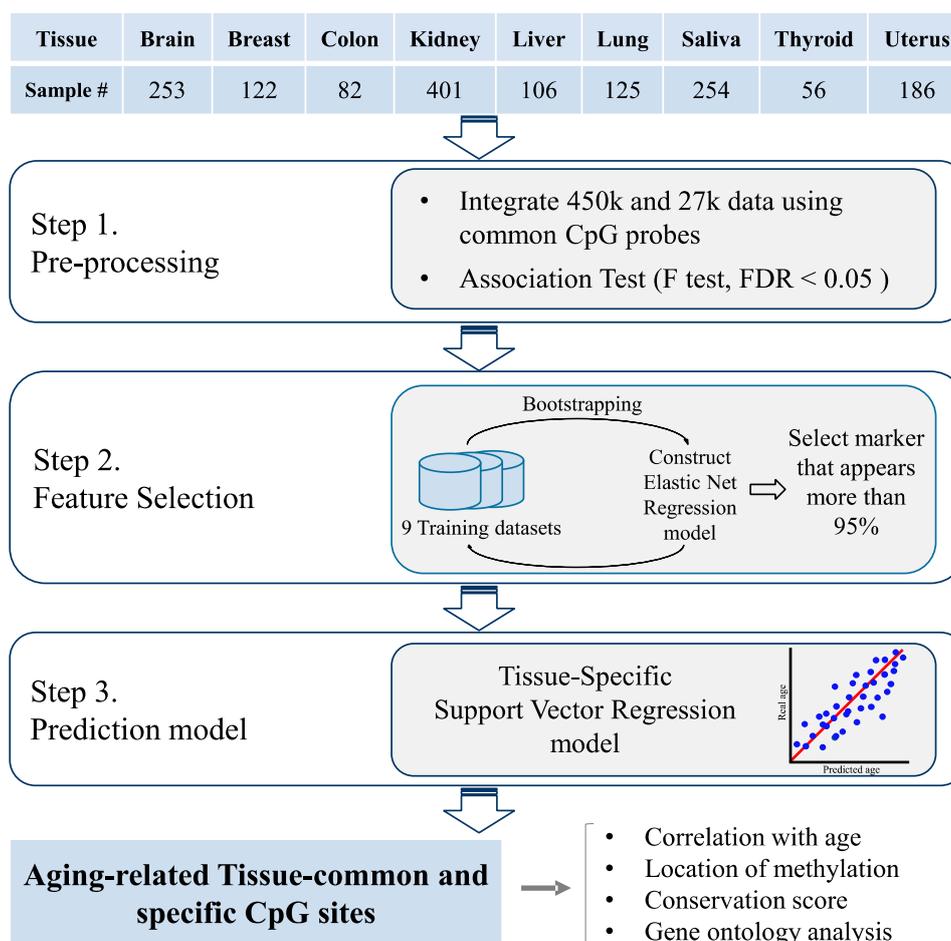
As aging-related methylation pattern research has progressed, some studies have identified epigenetic markers that accurately predict the chronological age of healthy people. Bocklandt et al. identified an age predictor using DNA methylation patterns derived from saliva samples [13]. Hannum et al. identified a more accurate age predictor using blood methylation data and reported 71 age-related Cytosine-phosphate-Guanine (CpG) sites known as “Hannum’s clock” [14]. Similarly,

a study showed the accurate prediction of age using only three CpG markers in DNA derived from blood [15]. It is important to note that these studies only used DNA from single tissues and primarily focused on blood [16]. On the other hand, Hovarth et al. reported a multi-tissue age predictor that could be applied to all tissues and cell types [17]. This study revealed 353 age-related CpG sites known as “Hovarth’s clock”. Due to its high performance in almost all tissue types, Hovarth’s clock has been widely used as a standard for estimating biological age using reliable epigenetic markers [18,19]. Since the epigenetic clock has emerged, many studies are still underway for methylation-based human aging prediction based on various machine learning techniques. More recently, Li et al. applied a gradient boosting approach to methylation age predictors, which provided even better accuracy for predicting human age [20]. However, these studies did not consider the tissue-specific characteristics of DNA methylation and only revealed CpG sites that explained common aging-related methylation patterns across all tissue types.

Age-related DNA–methylation shows distinct patterns in different tissues [21]. A comparison of methylation in adipose and liver tissues from young and old mice revealed that the age-related methylation sites in these two tissues showed distinctly different patterns [22]. Furthermore, tissue-common and tissue-specific methylation sites that are significantly correlated with age clearly show different characteristics [23]. These studies indicate a need for new epigenetic age estimators reflecting the specific characteristics of different tissue types.

Thus, to address the tissue-specific characteristics of DNA methylation changes corresponding to the aging process, in this study, we constructed tissue-specific chronological age predictors using DNA from the same tissues (Figure 1). Here, we hypothesized that DNA methylation pattern may be changed in different manners per each tissue over the aging process. Using DNA methylation data from nine distinct tissues, we constructed tissue-specific age predictors with different epigenetic markers. The validity of the suggested epigenetic markers is tested on the independent test dataset.

Finally, we identified new aging-related epigenetic markers that exhibit tissue-specific methylation patterns and methylation changes that were common across various tissue types. We further compared the characteristics of tissue-common and tissue-specific aging-related methylation patterns and showed that these patterns clearly differed.



**Figure 1.** Overview. We collected DNA methylation data from nine healthy human tissues from The Cancer Genome Atlas (TCGA) and Gene Expression Omnibus (GEO) datasets. We used both the 450 K and 27 K platforms to obtain as much data as possible and integrated the two platforms using common Cytosine-phosphate-Guanine (CpG) probes. To select informative features from the high-dimensional array data, initially we filtered the aging-related probes using an F test. Then, we performed feature selection using an elastic net regression and bootstrap analysis. Finally, we constructed tissue-specific support vector regression models for each of the nine tissues. Additionally, we analyzed the selected aging-related features by separating the data into two groups. We compared the characteristics of the CpG probes from the two groups, including the correlation with age, the methylation location, the conservation score, and gene ontology.

## 2. Materials and Methods

### 2.1. Tissue-Specific Methylation Datasets

All data were collected from publicly available databases, including The Cancer Genome Atlas (TCGA) [24] and the Gene Expression Omnibus (GEO) [25]. Tissue types with more than 30 normal samples were selected.

For the training dataset, most of the datasets used were collected from normal tissues in TCGA. The brain dataset consists of cerebellum, frontal lobe, cortex, and pons tissues from 150 subjects [26]. The breast dataset consists of normal tissue from TCGA (Breast invasive carcinoma: BRCA data). The colon dataset consists of adjacent normal tissue from TCGA (Colon adenocarcinoma: COAD and Rectum adenocarcinoma: READ data). The kidney dataset consists of adjacent normal tissue from TCGA (Kidney renal papillary cell carcinoma: KIRP and Kidney renal clear cell carcinoma:

KIRC data). The liver dataset consists of adjacent normal tissue from 62 Taiwanese Hepatocellular carcinoma (HCC) patients [27] and adjacent normal tissue from TCGA (Liver hepatocellular carcinoma: LIHC data). The lung dataset consists of adjacent normal tissue from TCGA (Lung adenocarcinoma: LUAD and Lung squamous cell carcinoma: LUSC data). The saliva dataset consists of samples from Khomani San individuals living in the South African Kalahari [28] and 197 alcohol-dependent subjects [29]. The thyroid dataset consists of adjacent normal tissue from TCGA (Thyroid carcinoma: THCA data). The uterus dataset consists of adjacent normal tissue from TCGA (Uterine Corpus Endometrial Carcinoma: UCEC data) and 152 normal uterine cervix samples [30]. Detailed information is summarized in Table 1.

**Table 1.** Training dataset. To train the model, we used DNA methylation data from healthy or normal adjacent samples obtained from nine different tissues. We collected methylation array data obtained using both the 27 K and 450 K platforms from TCGA and GEO datasets.

Data Set	Tissue Type	The Number of Patients		Age Range	Platform
GSE15745 [26]	Brain	253		16–95	HumanMethylation 27 K
TCGA [24]	Breast	95	122	28–90	HumanMethylation 450 K
TCGA [24]	Colon	27	82	35–88	HumanMethylation 27 K
TCGA [24]	Kidney	45	401	40–90	HumanMethylation 450 K
TCGA [24]	Liver	37	106	43–90	HumanMethylation 27 K
TCGA [24]	Lung	205	125	31–90	HumanMethylation 450 K
GSE37988 [27]	Saliva	196	254	33–86	HumanMethylation 27 K
TCGA [24]	Thyroid	49	56	20–81	HumanMethylation 450 K
GSE99029 [28]	Uterus	57	186	20–79	HumanMethylation 27 K
GSE34035 [29]		74		40–86	HumanMethylation 450 K
TCGA [24]		51		51–83	HumanMethylation 27 K
TCGA [24]		57		21–91	HumanMethylation 27 K
GSE30758 [30]		197		21–55	HumanMethylation 27 K
		56		15–81	HumanMethylation 450 K
		34		36–90	HumanMethylation 450 K
		152		19–55	HumanMethylation 27 K

The independent dataset was derived from normal adjacent tissues from the GEO. The brain dataset consists of 78 samples from the Brodmann area of the cerebral cortex [23]. The breast dataset consists of 23 non-neoplastic breast tissue samples [30] and adjacent normal tissues from patients with invasive breast cancer [31]. The colon dataset consists of 11 normal adjacent colon samples [32] and 87 normal colon biopsies used to study normal aging in the colonic mucosa [33]. The kidney dataset consists of 83 samples of normal adjacent resected kidney tissue [23]. The liver dataset consists of 79 normal liver samples from Germany [34] and control samples from a non-alcoholic fatty liver disease study [35]. The lung dataset consists of normal tissue obtained from smokers from a lung adenocarcinoma study [36]. The saliva dataset consists of saliva samples from an aging study collected from 54 males aged 18 to 73 years [37] and samples from identical twins aged 21 to 55 years [13]. The thyroid dataset consists of normal samples from a methylation study collected from 5 different normal and cancer tissue samples [38]. The uterus dataset was obtained from 152 women involved in a nested prospective case-control study [30]. Detailed information is summarized in Supplementary Table S1.

To generate coherent methylation features, all tissue-specific methylation datasets were selectively collected based on the use of either the Illumina 27 K or Illumina 450 K methylation chip as the assay platform. The Illumina methylation platform is an array-based quantification method for methylation levels at specific loci within the genome. Probes on the 27 K array target regions of the human genome to measure methylation levels at 27,578 CpG dinucleotides in 14,495 genes [39]. The Infinium HumanMethylation450 BeadChip array targets more than 450,000 CpG methylation sites within and outside of CpG islands [40]. The subsequent analyses used CpG features that were commonly measured in both the 27 K and 450 K platforms.

## 2.2. Association Test

To construct the age predictors, CpG probes that commonly appeared in the 27 K and 450 K platforms with less than 10 missing values were used. To combine two different datasets for training, such as in the liver tissue datasets, the 27 K and 450 K datasets were normalized. We applied quantile normalization using the “quantile\_transform” function in the preprocessing module provided by “Scikit-learn 0.19.1” [41]. Since more than 20,000 probes were too many to construct the aging models, a linear regression model was constructed for each probe. We performed *F*-tests between the beta values of each CpG probe and age and selected CpG sites with false discovery rate (FDR, Benjamini–Hochberg procedure [42]) values less than 0.05. Here, the *F* value of the regression analysis is the test result of the null hypothesis that the regression coefficients are all zero. Thus, CpG sites that passed the *F*-test (have a significant FDR of less than 0.05) represented CpG features that could have chronological age prediction power [43]. For this *F*-test, we used the “f\_regression” function in the feature-selection module provided by “Scikit-learn 0.19.1” [41].

## 2.3. Aging Model Construction

The feature selection process was performed using an elastic net algorithm [44], which is a multivariate linear model implemented in Python 3.5 and the package “Scikit-learn 0.19.1” [41]. The elastic net algorithm uses a penalized regression method that combines the lasso and ridge regression methods. This algorithm is ideal for high-dimensional data where the number of features is greater than the number of samples. We employed a bootstrap analysis by sampling the dataset with replacement and built elastic net regression models for each bootstrap cohort. We repeated this process 500 times and included markers that were present in more than 95% of all bootstrapping processes. Using these selected features, we constructed the final model using a support vector regression (SVR) algorithm. SVR is a regression method that maintains all the main features and characteristics of a support vector machine (SVM) [45]. SVR aims to find a regression function that minimizes the error and has a deviation from the target that is at most  $\epsilon$  at all times. To train models using SVR model, we should select several parameters like epsilon, degree of regularization, and kernel functions like linear kernel or RBF kernel. Every optimal parameter was estimated via 10-fold cross-validation by grid search [46–48].

## 2.4. Performance Comparison with Multi-Tissue Predictors

To validate the performance of the proposed models, we compared the models with two epigenetic age predictors. The first model is a multi-tissue age predictor that uses 353 CpG markers applicable to each tissue type [17]. For the comparison, we used 3931 methylation samples from 22 tissues and the R function for training described in this study. The second model was a multi-tissue model with SVR. We used the SVR algorithm to predict age using selected features. To validate that the selected features explained the aging process in each tissue, we applied the SVR algorithm in the multi-tissue age predictors. We collected the dataset specified by the study. Using 353 CpG markers as multi-tissue age predictors, we applied the SVR algorithm with the linear kernel. For model construction, parameters such as epsilon and C were selected by 10-fold cross-validation; the same procedure was applied for the tissue-specific age predictors.

To evaluate model performance, we used three regression metrics. The first metric is a root-mean-square error (RMSE). The RMSE represents the square root of the second sample moment of the differences between predicted and observed values. The second metric is the mean absolute deviation (MAD). The MAD represents the average vertical distance between each point and the  $Y = X$  line. The third metric is Pearson’s correlation coefficient, which is also referred to as Pearson’s  $r$ . The  $r$  value represents the measure of the linear correlation between predicted and observed values and has a value between +1 and –1, where 1 is a total positive linear correlation, 0 is no linear correlation, and –1 is a total negative linear correlation.

For comparison with the multi-tissue age predictor using other regression metrics, we used two types of metrics. The first metric is the mean absolute percentage error (MAPE). The MAPE is a measure of the prediction accuracy of a forecasting method in statistics. The absolute value in this calculation is summed for every forecasted point in time and divided by the number of fitted points  $n$ . Multiplying by 100% makes the value a percentage error [49]. The other metric is Theil's U statistic. This metric is a relatively accurate measure that compares the forecasted results with the results of forecasting with minimal historical data [50].

### 2.5. Conservation Score Calculation

We examined the phastCons score of the CpG sites within the genomic regions provided by the human methylation 450 K annotation file [51]. PhastCons score is a probability that each nucleotide belongs to a conserved element through the evolution process. In this case, we used Phastcons score for multi alignment to 99 vertebrate genomes to the human genome. We calculated the average phastCons score for each base of the genomic regions.

### 2.6. GO Analysis

The GO analysis was performed using the ClueGO plugin in Cytoscape 3.5.1 [52]. We used the GO biological process database as a reference. Gene lists were selected using the closest gene to each CpG marker that was provided by the Illumina human methylation 27 K platform annotation file.

## 3. Results

### 3.1. Tissue-Specific Age Prediction Model

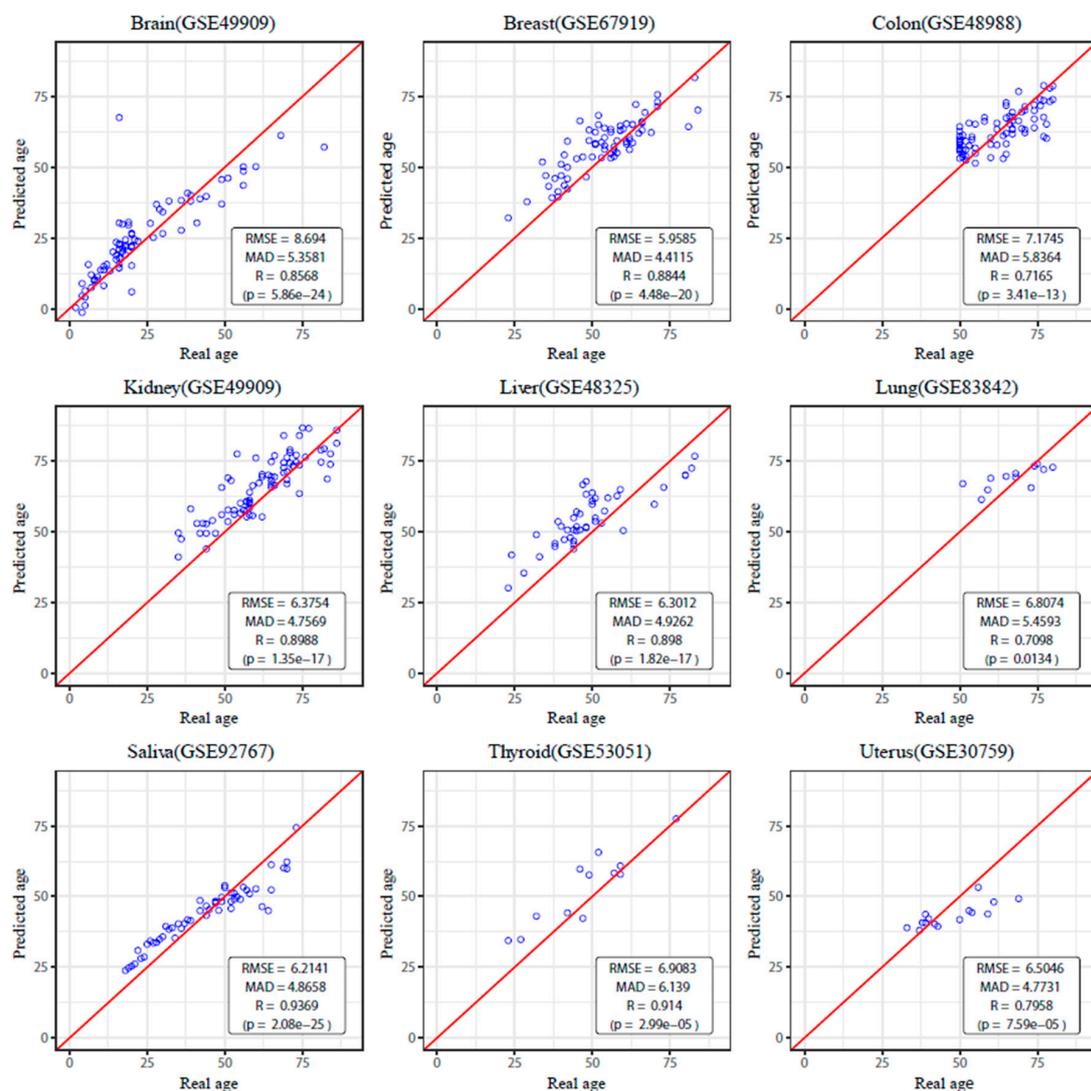
In this study, we used 21,396 probes that could be measured using the Illumina Infinium II assay and were common to both the 27 K and 450 K platforms.

Initially, the number of CpG features was more than 100 times greater than the average number of samples. Therefore, to filter out features that are not related to aging, we performed association tests between the  $\beta$ -values of each CpG site and age. As a result, we selected CpG sites that had a false discovery rate (FDR)  $< 0.05$ . After the association test, the average number of features of each tissue was reduced to 8800. Then, we selected features to be used in the age prediction model by using a penalized multivariate regression method (Table 2). Additionally, we counted the number of occurrence in features from each tissue-specific age predictors. cg22736354 methylation marker is found as common marker of all nine tissue-specific age predictors. Detailed information is provided in Supplementary Table S3. Finally, using these selected features, we constructed nine tissue-specific age prediction models using a SVR algorithm (Figure 1). We used the SVR algorithm in the final regression modeling because it showed lower mean absolute deviation (MAD) values than the elastic net regression in almost all tissue types. The results are presented in Supplementary Figure S1. In addition, the SVR method had better epigenetic age prediction performance than the multivariate regression method [53]. In previous epigenetic aging studies, elastic net regressions have been typically used for both feature selection and age prediction [14–17]. As methylation array data are not abundant in normal samples, use of the SVR algorithm instead of a simple multivariate linear regression appeared to perform well in analyses of the high-dimensional bioinformatic data.

**Table 2.** The number of features of each tissue. This table lists the number of final features used for model construction. We used these features for the subsequent feature analysis.

	Brain	Breast	Colon	Kidney	Liver	Lung	Saliva	Thyroid	Uterus	Total
Total	256	249	86	371	248	148	280	221	46	1460
Common	93	92	17	153	53	49	103	99	35	247
Specific	163	157	69	218	195	99	177	122	11	1213
Positive	170	202	54	281	186	67	232	151	35	989
Negative	86	47	32	90	62	81	48	70	11	471

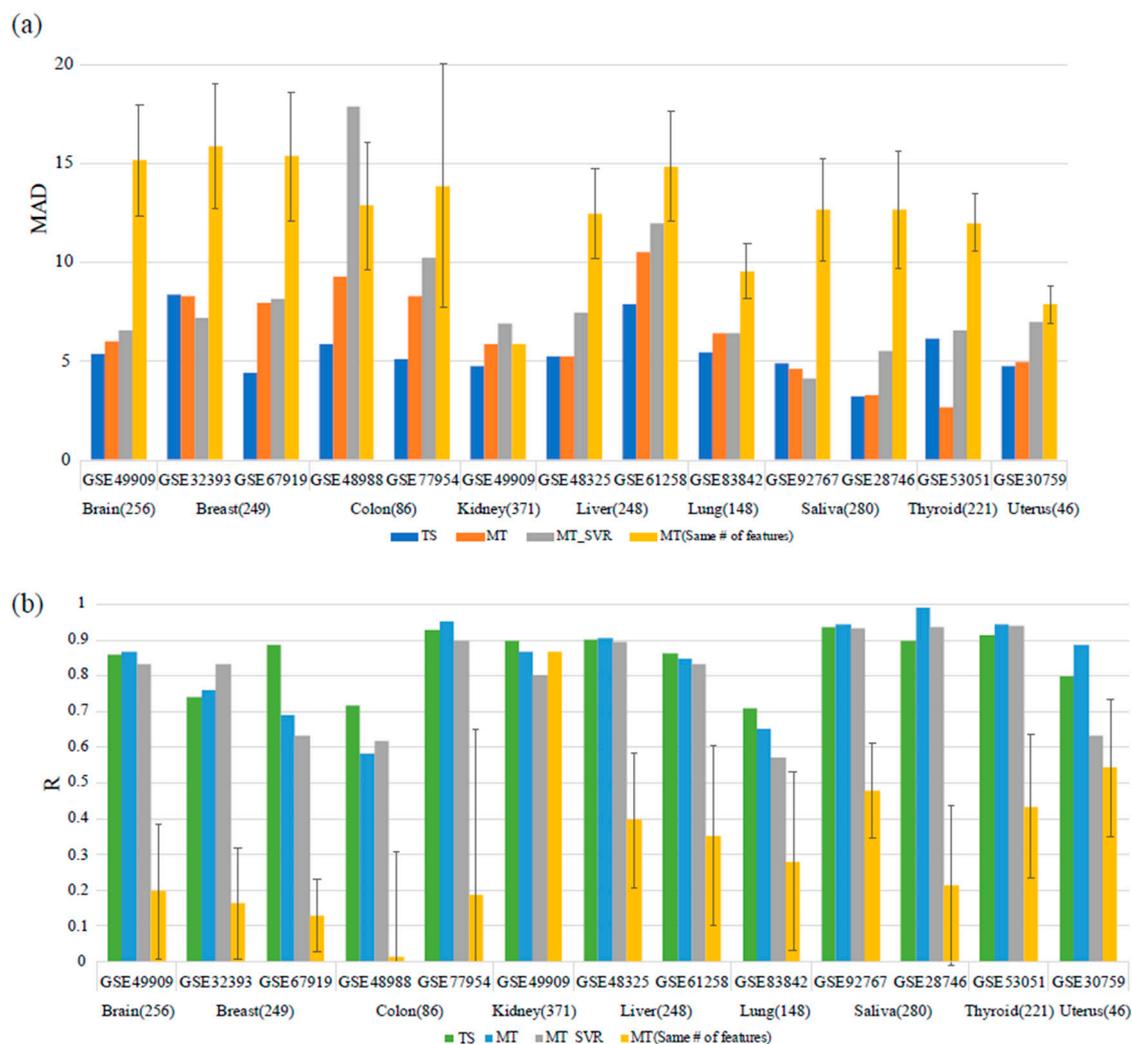
We validated the performance of the tissue-specific age predictors using several metrics, including the root mean squared error (RMSE), mean absolute deviation (MAD), and Pearson's correlation coefficient ( $r$ ), to compare the predicted age to the chronological age. According to these accuracy measures, the tissue-specific age predictors performed well in all nine tissue types. We performed 10-fold cross-validation and external validation to verify that the model was well trained.  $K$ -fold cross-validation is a model validation technique that uses a partitioned original sample to test a model trained with the remaining  $k-1$  subsamples. We set the parameter  $k$  as 10 [47]. For the 10-fold cross-validation, the RMSE was 4.679 (years), MAD was 3.519 (years), and  $r$  was 0.9619 on average for the nine tissue models (Supplementary Figure S2). To test that nine tissue-specific age predictors performed well in real-world methylation data, we also validated the model in an independent test dataset for every nine tissues from GEO database. Regarding the external validation, the RMSE was 6.512 (years), MAD was 5.119 (years), and  $r$  was 0.845 on average (Figure 2).



**Figure 2.** External Validation. We tested the tissue-specific age predictors listed in an independent dataset and found that the mean absolute deviation (MAD) was 5.119 years and the correlation coefficient was 0.845 on average for the nine models. The results in the brain (MAD = 5.358,  $r = 0.859$ ), breast (MAD = 4.412,  $r = 0.884$ ), colon (MAD=5.836,  $r = 0.717$ ), kidney (MAD = 4.757,  $R=0.899$ ), liver (MAD = 4.926,  $R = 0.898$ ), lung (MAD = 5.459,  $R = 0.710$ ), saliva (MAD = 4.866,  $R = 0.937$ ), thyroid (MAD = 6.14,  $R = 0.910$ ), and uterus (MAD = 4.773,  $R = 0.796$ ) are shown.

### 3.2. Comparison with the Multi-Tissue Age Predictor

To validate the tissue-specific age predictors, we compared the performance of two models. The first model was the multi-tissue age prediction model described by Hovarth, which used 353 aging-related CpG markers [17]. The second model was a multi-tissue age predictor that applied the SVR algorithm. In our results, generally, the tissue-specific age predictors had lower MAD values than the multi-tissue age predictors (Figure 3). For comparison, we used the same test dataset that we used for validating the proposed age predictors (see Methods 2.1 for details). Also these test datasets were not used in the training process of multi-tissue age predictor, which were also appropriate for comparing the performance of multi-tissue age predictor (Supplementary Table S2).



**Figure 3.** Comparison with a previous study. The results were compared between the proposed tissue-specific age predictors and multi-tissue age predictors for an independent dataset collected in this study. For the abbreviations, TS means Tissue-specific age predictor we suggested, MT means Multi-tissue age predictor and MT\_SVR is Support Vector Regression Model constructed by using features of multi-tissue age predictor for equitable comparison. MT(same # of features) means a Multi-tissue age predictor using the same number of features in TS model. The numbers in parentheses refer to the number of markers of each tissue-specific model. (a) When comparing MAD values, the tissue-specific age predictors showed better performance except for the breast and thyroid tissue datasets. (b) When comparing the  $R$  values as coefficient of correlation, some tissues showed better performance in the multi-tissue age predictors, but the results were similar.

Although the  $r$  values of the multi-tissue age predictors were slightly higher than those of the tissue-specific age predictors in some tissues, the models exhibited similar results. We compared the models using other metrics that are commonly used in regression problems, such as the mean absolute percentage error (MAPE) and Theil's  $U$  statistics, and found that the tissue-specific age predictors tended to show better results (Supplementary Figure S3). As these improved results may be due to the superiority of SVR and not the explanatory power of the selected features, we applied the SVR algorithm to the dataset and features described by Hovarth. As a result, the proposed tissue-specific age predictors generally showed better performance than the multi-tissue age predictors with the SVR model. This result shows that tissue-specific age predictors can reflect the aging process well despite the smaller numbers of training data and features. For some dataset like saliva (GSE92676), or thyroid (GSE53051) multi-tissue age predictor showed better results in prediction error. However, tissue-specific age predictors showed better or comparable results using a smaller number of training samples and features. Considering that multi-tissue age predictors used about 4000 samples for training the model, which also added plentiful blood dataset samples, tissue-specific age predictors used only about 200 samples for each tissue on average for training and used a smaller number of features to predict the age of the samples. Despite the unfair condition, tissue-specific age predictors predicted age of samples with comparable level with multi-tissue age predictors by also considering the tissue-specific properties. We also compared how many selected features overlap with Hovarth's clock (353 CpGs). Methylation sites used for features of multi-tissue age predictors more belonged to tissue-common group (Supplementary Figure S4).

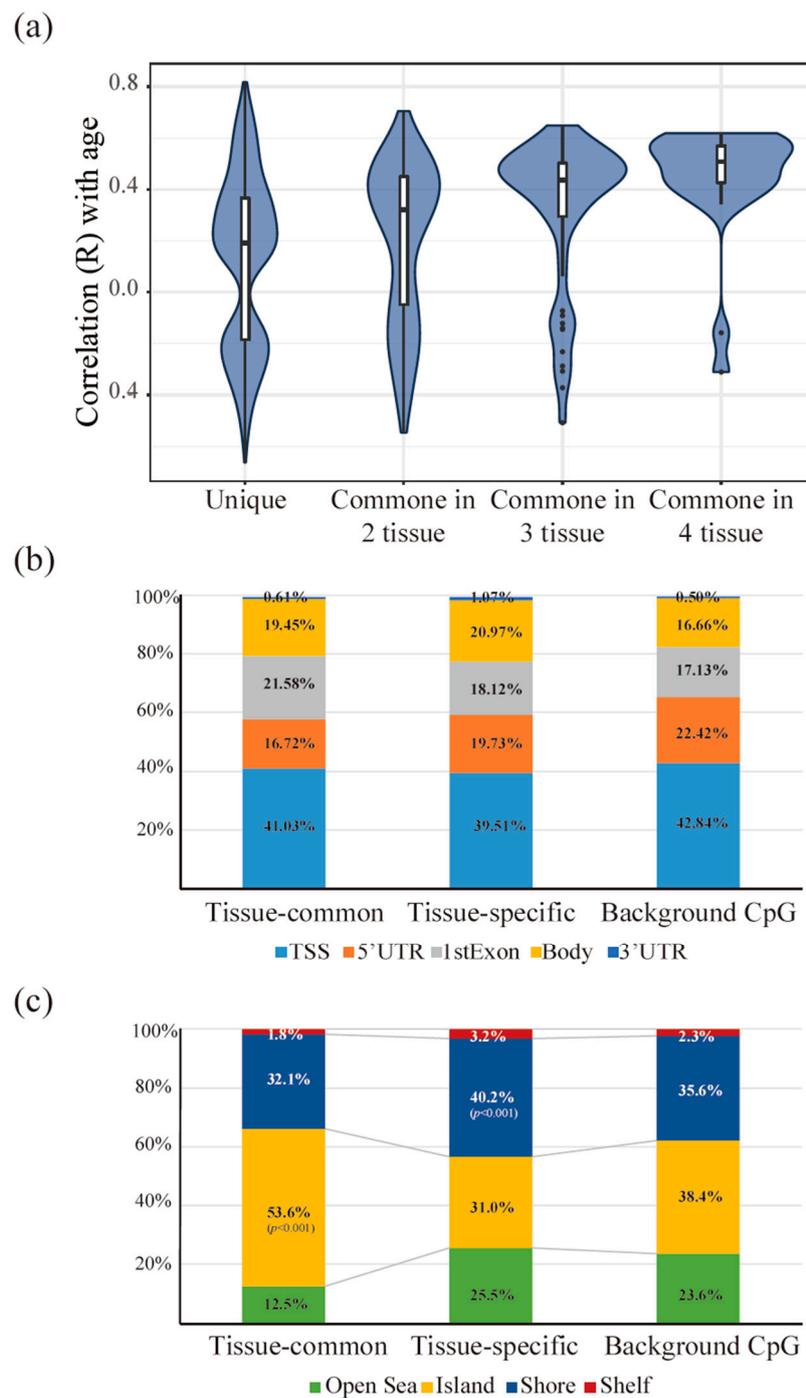
In addition, our results suggest that age-related DNA methylation changes have both tissue-specific and tissue-common patterns, and we have selected the features that reflect both of these characteristics.

To determine whether the tissue-common and tissue-specific patterns had different characteristics, we divided the selected features into two groups and analyzed their features. We defined the "tissue-common group" as CpG sites that were commonly used in more than two tissue models. In contrast, the "tissue-specific group" was defined as CpG sites that were uniquely used in only one tissue model. In total, we found 247 tissue-common aging markers and 1,213 tissue-specific aging markers. The number of aging markers in each tissue type is listed in Table 2.

Several studies have shown aging-related methylation patterns in blood. Thus, we aimed to validate that the tissue-common group was consistent with the findings from these previous studies. In a review on DNA methylation and healthy human aging, the authors filtered 11 CpG sites associated with age in the blood that appeared in at least four of the eight studies [54]. Nine of the 11 CpG sites in our derived tissue-common group were congruent with aging. Furthermore, compared with Hovarth's clock, 353 CpG sites were included in more of the tissue-common features than in the tissue-specific features. These results show that the tissue-common markers in our models reflect epigenetic aging patterns that generally appear throughout various tissues.

### 3.3. Tissue-Common and Tissue-Specific Features of Age-Dependent Methylation in Tissues

We hypothesized that tissue-specific aging signatures would show methylation patterns that differed from those commonly observed. To validate this assumption, we compared the tissue-common methylation patterns with the tissue-specific methylation patterns. First, we compared the direction of methylation that accompanied the aging process. We defined "positive ageCGs" as CpG sites whose methylation increased during the aging process and "negative ageCGs" as CpG sites whose methylation decreased during the aging process. Generally, more positive ageCGs were found, but the tissue-specific group had more negative ageCGs than the tissue-common group (Supplementary Figure S5). Interestingly, as the number of markers that appeared in the tissue types increased, the ratio of positive ageCGs also increased (Figure 4a). This finding could suggest that general methylation patterns increase with age; however, decreasing methylation patterns that were associated with tissue-specific aging were also observed.

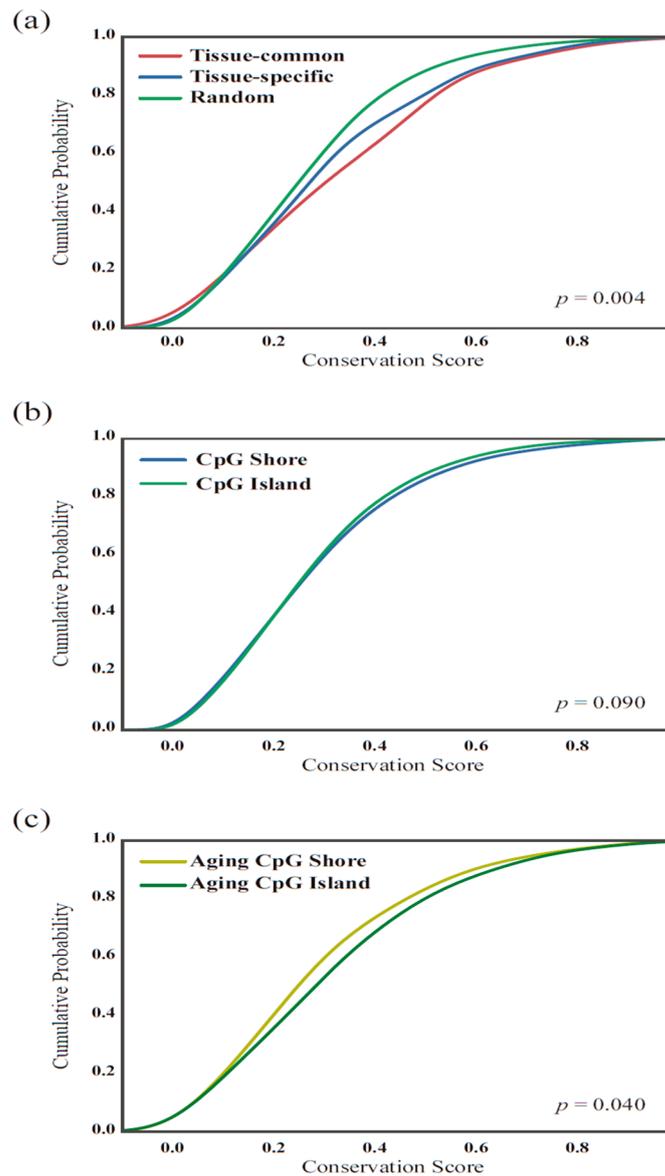


**Figure 4.** Different patterns between the tissue-common and tissue-specific groups. Different characteristics of the methylation direction and location between the tissue-common and tissue-specific groups (a–c) (a) When the correlation coefficient of the beta values of each methylation site and age were compared, more positive ageCGs were found in aging-related CpG sites. However, the ratio of negative ageCGs was higher in the unique groups defined as tissue-specific. In addition, the ratio of positive ageCGs increased as the number of appearances in the tissue model increased. (b) When the methylation locations based on the gene structure were compared, no significant changes were found between the two groups ( $p = 0.4059$ ). We compared the methylation location sites in the 27 K platform and found more gene body regions that contained aging-related CpG sites ( $p < 10^{-6}$ ). (c) We compared the locations of the methylation sites in the CpG islands and CpG shores and found more CpG island regions in the tissue-common group ( $p < 10^{-6}$ ) and more CpG shore regions in the tissue-specific group ( $p = 0.0001$ ).

Also, to compare the general methylation patterns that emerged during aging, we divided the data into a younger group and an older group. For the training data, we selected the five oldest and five youngest samples from each tissue type. For each tissue, the age difference between the two groups was greater than 40 years. No significant difference was observed in the total 27 K methylation patterns between the two groups. However, by comparing the methylation patterns of the selected aging-related features, we observed that the older group tended to be more highly methylated than the younger group in each tissue type except for the lung tissue (Supplementary Figure S6). This result is consistent with the abundance of positive ageCGs found in aging-related methylation sites.

Next, we compared the ratio of the gene structures in which each CpG marker was located. The distribution of the locations in the gene structures in both the common and unique CpG groups did not significantly differ (Fisher's exact test,  $p = 0.4059$ ). However, the aging-related methylation markers were more likely to be located in the gene bodies than the location distribution of all CpG sites in the 27 K platform (Fisher's exact test,  $p < 10^{-6}$ ) (Figure 4b). Although the result was not significant, the tissue-common group included more exon regions than the tissue-specific group (Fisher's exact test,  $p = 0.1303$ ) (Supplementary Figure S7). In addition, we compared the ratios of the CpG islands to the CpG shore regions in which each CpG marker was located. The tissue-common group tended to be located in the CpG island regions, whereas the tissue-specific group tended to be located in the CpG shore regions (Fisher's exact test,  $p < 10^{-6}$ ,  $p = 0.0001$ ) (Figure 4c). These results are consistent with results from a previous study comparing age-related tissue-common and tissue-specific patterns in four tissues [23]. The results show that tissue-specific and tissue-common methylation patterns appear in different locations and that each tissue type may have distinct mechanisms that operate during the normal aging process.

Furthermore, we assumed that the methylation patterns appearing in various tissues might be more evolutionarily conserved than the patterns from tissue-specific mechanisms. To test this assumption, we calculated the average evolutionary conservation (phastCons) score of the chromosome region of each CpG marker from the Illumina annotation file (see Methods for details). As expected, the distribution of the conservation scores of the tissue-common and tissue-unique groups differed (Kolmogorov-Smirnov test,  $p = 0.00413$ ). The aging-related CpG markers tended to be located in more conserved regions than randomly selected CpG sites (Figure 5a). This result may be due to the original conservation score differences in the CpG island regions. Thus, we compared the conservation score distributions of the whole CpG island and shore regions provided by the Illumina 27 K annotation file but found few differences ( $p = 0.09$ ) (Figure 5b). However, when we compared the conservation score distributions of the CpG islands and shore regions of the aging markers, we found significant differences ( $p = 0.04$ ) (Figure 5c). We thought that this conservation score difference was not due to the known evolutionary conserveness difference in CpG Island and Shore region. However, it is because of tissue-specific and common methylation differences throughout the aging process.



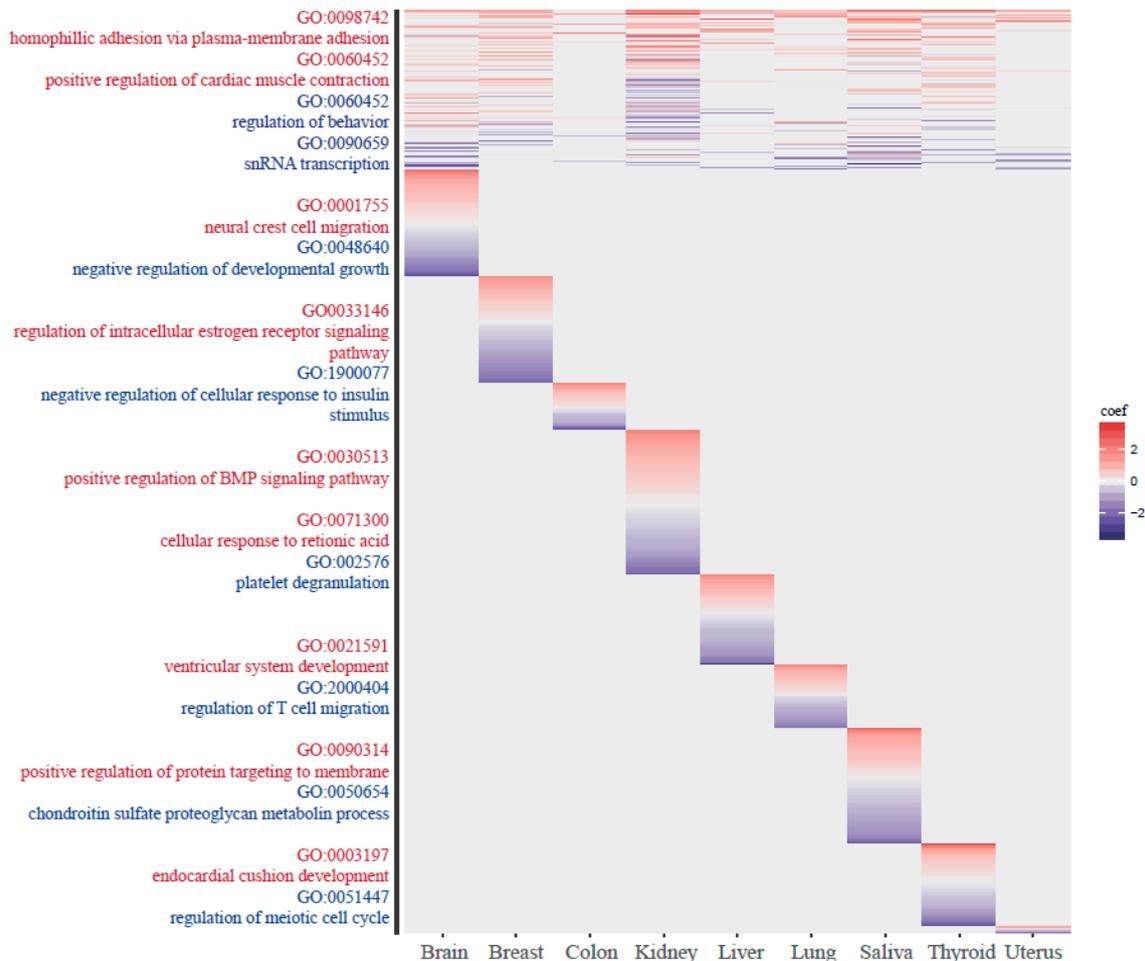
**Figure 5.** Conservation scores. (a) Comparison of the cumulative distribution of the average conservation scores between the tissue-common and tissue-specific groups. The two distributions of the tissue-common and tissue-specific group show significant differences using the Kolmogorov-Smirnov test ( $p = 0.004$ ). (b) As the two groups showed differences in the methylation locations based on CpG islands, we compared the conservation score of each CpG site in the CpG island and shore regions from the 27 K methylation platform but did not find a significant difference ( $p = 0.09$ ). (c) However, the conservation scores of the aging-related CpG island and shore regions slightly differed ( $p = 0.04$ ).

### 3.4. The Functionality of Tissue-Common and Specific Methylation Regions

Finally, we performed a gene ontology (GO) analysis to validate the function of the closest gene to the selected markers. We assumed that the selected methylation markers affected the function of the gene closest to the site. To analyze the gene ontology, the markers were divided into the positive ageCG and negative ageCG groups, since we suspected that the functions of the positive ageCGs and negative ageCGs might differ.

In the tissue-common group, GO terms such as cell-cell adhesion and cardiac muscle contraction appeared in the positive direction, and GO terms such as regulation of behavior and snRNA transcription appeared in the negative direction. In the tissue-specific group with positive ageCGs, tissue-specific GO

terms were associated with brain and breast tissue, including neural crest cell migration and mammary gland lobule development. In the tissue-specific group with negative ageCGs, immune-related terms were found, such as regulation of T cell migration. The number of T cells and the distribution of lymphocyte subsets have been shown to vary with age [55]. The GO terms for each group are shown in Figure 6, and the top lists of GO terms for tissue-common and tissue-specific group are provided in Supplementary Tables S4 and S5.



**Figure 6.** Gene ontology (GO) analysis. GO terms for the closest gene of each methylation group are shown. Because we conducted the gene ontology analysis by separating the sites into positive and negative ageCGs, red color indicates the GO terms from the positive ageCGs and blue color indicates the negative ageCGs.

#### 4. Discussion

Although many studies have investigated aging-related epigenetic markers using methylation arrays, these studies are limited by only considering the common characteristics of aging across tissues. In this study, we constructed nine tissue-specific age prediction models to study the tissue specificity of DNA methylation. Using a penalized regression model, reliable features that affect aging were selected. We selected fewer than 400 epigenetic markers for each tissue that may reflect the aging processes of that tissue. The number of selected features differed for each tissue, and these differences may reflect the different aging processes of each tissue. For instance, tissue models with more CpG features may suggest that the tissues have undergone more biological changes during the aging process. However, these differences could also be the result of the different numbers of features derived from the model training process.

Generally, the tissue-specific age predictors showed better performance than the multi-tissue predictors despite the smaller number of features and the smaller training dataset. However, obtaining better results in the comparison does not necessarily mean that the tissue-specific age predictors represent a better biological age prediction method. The multi-tissue age predictors estimated the chronological age of each tissue with high accuracy. However, this model was limited because the specificity of DNA methylation patterns in different tissues was not considered. Moreover, tissue-specific age predictors have an advantage in providing tissue-specific biomarkers. By constructing a model for each tissue type, we found different aging markers that did not appear in other tissues. This method was limited because the feature selection was time-consuming. Moreover, because this method was applied to only nine tissues, finding aging-related methylation markers in other tissues could prove to be difficult.

Although the two models have pros and cons, the better performance of the tissue-specific age predictors suggests that aging-related DNA methylation patterns have both tissue-specific and tissue-common characteristics. Therefore, we compared the tissue-common and tissue-specific aging-related DNA methylation patterns to determine whether the two groups showed different characteristics. First, more positive ageCGs were positively correlated with age, suggesting that generally aging-related methylation regions increased throughout an individual's lifetime. Interestingly, the tissue-specific groups had more negative ageCGs than positive ageCGs; this phenomenon showed that decreasing methylation patterns also existed in tissue-specific aging mechanisms. In addition, these two groups are located in remarkably different regions based on their proximity to CpG islands. The tissue-common methylation regions involved in aging tended to be more conserved than the tissue-specific regions. This result suggests that the tissue-specific and tissue-common groups show different patterns because biologically important regions are more conserved throughout evolution.

Finally, the two groups showed different functionalities. To determine the differences in each methylation group, we performed a GO analysis by separating the data into positive and negative ageCGs. Interestingly, the genes in regions of both positive and negative tissue-common groups tended to show GO terms such as muscle contraction and behavior. Furthermore, each tissue showed different functions, although not all tissues showed the same tissue-specific ontology terms. Although tissue-specific terms typically appeared in the negative ageCGs, tissue-specific terms in the positive ageCGs were found from brain and breast tissue. We believe that this finding is the result of the small number of negative ageCGs; thus, many tissue-specific terms were not statistically significant. In addition, determining which tissue is aging at the fastest rate is an important question in aging studies. Some studies have investigated different aging rates in specific tissues [56,57]. However, comparing different biological ages predicted by age predictors of each tissue is inappropriate, because the prediction may be the result of the bad performance of the model. We did not directly address the rate of aging between tissues. We investigated that the degrees of the slope between age and methylation level of cg22736354, which is one of the tissue-common markers. According to the aging process, we noted that the slopes of the methylation level of this single cg22736354 were different for each tissue type. We confirmed that the slope of colon tissue is smallest, and in the brain tissue, the slope value is higher than other tissues. These results suggest that aging marker in colon tissue more slowly than other tissues (supplementary Table S6).

## 5. Conclusions

This study revealed epigenetic age predictors that reflect the characteristics of DNA methylation and tissue-specific aging mechanisms. The tissue-common markers found in this study may be used as reliable epigenetic aging markers that reflect aging-related DNA methylation changes that commonly appear in many tissues. Furthermore, these tissue-specific markers may be used as targets to identify tissue-specific epigenetic aging signatures. However, the cellular mechanisms of DNA methylation require further investigation. Since a clear description of the relationship between DNA methylation

and aging is lacking, each tissue type may have different biological mechanisms, and thus, both tissue-common and tissue-specific aging mechanisms need to be considered in future studies.

Although the proposed tissue-specific age-prediction model showed remarkable performance, the method still has room for improvement. First, with the development of next-generation sequencing (NGS) technology, non-CpG methylation has been measured and observed, and the role of non-CpG methylation in specific cells has been interpreted to some extent in recent years. In this sense, non-CpG methylation can also be considered an aging-prediction marker [58]. Additionally, with the accumulation of datasets, more sophisticated machine learning models [59–61], such as deep neural networks (DNN) [62,63], can be applied to identify and interpret age-related features that cannot be captured with simple linear or non-linear regression models.

**Supplementary Materials:** The following are available online at <http://www.mdpi.com/2073-4425/10/11/888/s1>, Figure S1: Comparison of SVR and Elastic Net regression method; Figure S2: Internal validation; Figure S3: Comparison with other metric Comparison with other metric; Figure S4: Number of common features with Hovarth; Figure S5: Ratio of positive and negative ageCGS; Figure S6a: General methylation pattern difference in young and old group; Figure S6b: Aging-related methylation pattern difference in young and old group; Figure S7: Exon and intron ratio in tissue-common and tissue-specific group, Table S1: Independent dataset; Table S2: Common dataset with Hovarth et al; Table S3: Number of Common features in tissue-common CpG groups; Table S4: Gene Ontology Analysis of tissue-common group; Table S5a: Gene Ontology Analysis of positive ageCGs of tissue-specific group; Table S5b: Gene Ontology Analysis of negative ageCGs of tissue-specific group; Table S6: The comparison with cg22736354 methylation and age in each tissue, Supplementary\_Aging\_marker\_list.xlsx contains detailed CpG marker information.

**Author Contributions:** H.N. designed and conceived the study. H.C. implemented the study and drafted the manuscript. Both authors revised and approved the final manuscript. S.J. Formal analysis; Investigation; writing—review and editing.

**Funding:** This work was supported by the Bio-Synergy Research Project (NRF-2017M3A9C4092978) of the Ministry of Science, and was supported by the GIST Research Institute (GRI) in 2019.

**Conflicts of Interest:** The authors declare no competing interests.

## References

1. Lopez-Otin, C.; Blasco, M.A.; Partridge, L.; Serrano, M.; Kroemer, G. The hallmarks of aging. *Cell* **2013**, *153*, 1194–1217. [[CrossRef](#)]
2. West, M.D.; Sternberg, H.; Labat, I.; Janus, J.; Chapman, K.B.; Malik, N.N.; de Grey, A.D.; Larocca, D. Toward a unified theory of aging and regeneration. *Regen. Med.* **2019**, *14*, 867–886. [[CrossRef](#)]
3. Baker, G.T., 3rd; Sprott, R.L. Biomarkers of aging. *Exp. Gerontol.* **1988**, *23*, 223–239. [[CrossRef](#)]
4. Hudgins, A.D.; Tazearslan, C.; Tare, A.; Zhu, Y.; Huffman, D.; Suh, Y. Age- and tissue-specific expression of senescence biomarkers in mice. *Front. Genet.* **2018**, *9*, 59. [[CrossRef](#)] [[PubMed](#)]
5. Peters, M.J.; Joehanes, R.; Pilling, L.C.; Schurmann, C.; Conneely, K.N.; Powell, J.; Reinmaa, E.; Sutphin, G.L.; Zhernakova, A.; Schramm, K.; et al. The transcriptional landscape of age in human peripheral blood. *Nat. Commun.* **2015**, *6*, 8570. [[CrossRef](#)] [[PubMed](#)]
6. De Magalhaes, J.P.; Curado, J.; Church, G.M. Meta-analysis of age-related gene expression profiles identifies common signatures of aging. *Bioinformatics* **2009**, *25*, 875–881. [[CrossRef](#)] [[PubMed](#)]
7. Harley, C.B.; Futcher, A.B.; Greider, C.W. Telomeres shorten during ageing of human fibroblasts. *Nature* **1990**, *345*, 458–460. [[CrossRef](#)]
8. Heidinger, B.J.; Blount, J.D.; Boner, W.; Griffiths, K.; Metcalfe, N.B.; Monaghan, P. Telomere length in early life predicts lifespan. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 1743–1748. [[CrossRef](#)]
9. Blackburn, E.H.; Gall, J.G. A tandemly repeated sequence at the termini of the extrachromosomal ribosomal RNA genes in Tetrahymena. *J. Mol. Biol.* **1978**, *120*, 33–53. [[CrossRef](#)]
10. Jung, M.; Pfeifer, G.P. Aging and DNA methylation. *BMC Biol.* **2015**, *13*, 7. [[CrossRef](#)]
11. Teschendorff, A.E.; Menon, U.; Gentry-Maharaj, A.; Ramus, S.J.; Weisenberger, D.J.; Shen, H.; Campan, M.; Noushmehr, H.; Bell, C.G.; Maxwell, A.P.; et al. Age-dependent DNA methylation of genes that are suppressed in stem cells is a hallmark of cancer. *Genome Res.* **2010**, *20*, 440–446. [[CrossRef](#)] [[PubMed](#)]

12. Horvath, S.; Zhang, Y.; Langfelder, P.; Kahn, R.S.; Boks, M.P.; van Eijk, K.; van den Berg, L.H.; Ophoff, R.A. Aging effects on DNA methylation modules in human brain and blood tissue. *Genome Biol.* **2012**, *13*, R97. [[CrossRef](#)] [[PubMed](#)]
13. Bocklandt, S.; Lin, W.; Sehl, M.E.; Sanchez, F.J.; Sinsheimer, J.S.; Horvath, S.; Vilain, E. Epigenetic predictor of age. *PLoS ONE* **2011**, *6*, e14821. [[CrossRef](#)] [[PubMed](#)]
14. Hannum, G.; Guinney, J.; Zhao, L.; Zhang, L.; Hughes, G.; Sada, S.; Klotzle, B.; Bibikova, M.; Fan, J.B.; Gao, Y.; et al. Genome-wide methylation profiles reveal quantitative views of human aging rates. *Mol. Cell* **2013**, *49*, 359–367. [[CrossRef](#)] [[PubMed](#)]
15. Weidner, C.I.; Lin, Q.; Koch, C.M.; Eisele, L.; Beier, F.; Ziegler, P.; Bauerschlag, D.O.; Jockel, K.H.; Erbel, R.; Muhleisen, T.W.; et al. Aging of blood can be tracked by DNA methylation changes at just three CpG sites. *Genome Biol.* **2014**, *15*, R24. [[CrossRef](#)]
16. Marioni, R.E.; Shah, S.; McRae, A.F.; Chen, B.H.; Colicino, E.; Harris, S.E.; Gibson, J.; Henders, A.K.; Redmond, P.; Cox, S.R.; et al. DNA methylation age of blood predicts all-cause mortality in later life. *Genome Biol.* **2015**, *16*, 25. [[CrossRef](#)]
17. Horvath, S. DNA methylation age of human tissues and cell types. *Genome Biol.* **2013**, *14*, R115. [[CrossRef](#)]
18. Horvath, S.; Pirazzini, C.; Bacalini, M.G.; Gentilini, D.; Di Blasio, A.M.; Delledonne, M.; Mari, D.; Arosio, B.; Monti, D.; Passarino, G.; et al. Decreased epigenetic age of PBMCs from Italian semi-supercentenarians and their offspring. *Aging* **2015**, *7*, 1159–1170. [[CrossRef](#)]
19. Chen, B.H.; Marioni, R.E.; Colicino, E.; Peters, M.J.; Ward-Caviness, C.K.; Tsai, P.C.; Roetker, N.S.; Just, A.C.; Demerath, E.W.; Guan, W.; et al. DNA methylation-based measures of biological age: Meta-analysis predicting time to death. *Aging* **2016**, *8*, 1844–1865. [[CrossRef](#)]
20. Li, X.; Li, W.; Xu, Y. Human age prediction based on DNA methylation using a gradient boosting regressor. *Genes (Basel)* **2018**, *9*, 424. [[CrossRef](#)]
21. Jung, S.E.; Shin, K.J.; Lee, H.Y. DNA methylation-based age prediction from various tissues and body fluids. *BMB Rep.* **2017**, *50*, 546–553. [[CrossRef](#)] [[PubMed](#)]
22. Maegawa, S.; Hinkal, G.; Kim, H.S.; Shen, L.; Zhang, L.; Zhang, J.; Zhang, N.; Liang, S.; Donehower, L.A.; Issa, J.P. Widespread and tissue specific age-related DNA methylation changes in mice. *Genome Res.* **2010**, *20*, 332–340. [[CrossRef](#)] [[PubMed](#)]
23. Day, K.; Waite, L.L.; Thalacker-Mercer, A.; West, A.; Bamman, M.M.; Brooks, J.D.; Myers, R.M.; Absher, D. Differential DNA methylation with age displays both common and dynamic features across human tissues that are influenced by CpG landscape. *Genome Biol.* **2013**, *14*, R102. [[CrossRef](#)] [[PubMed](#)]
24. The Cancer Genome Atlas Research Network; Weinstein, J.N.; Collisson, E.A.; Mills, G.B.; Shaw, K.R.; Ozenberger, B.A.; Ellrott, K.; Shmulevich, I.; Sander, C.; Stuart, J.M. The cancer genome atlas pan-cancer analysis project. *Nat. Genet.* **2013**, *45*, 1113–1120. [[CrossRef](#)]
25. Edgar, R.; Domrachev, M.; Lash, A.E. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.* **2002**, *30*, 207–210. [[CrossRef](#)] [[PubMed](#)]
26. Gibbs, J.R.; van der Brug, M.P.; Hernandez, D.G.; Traynor, B.J.; Nalls, M.A.; Lai, S.L.; Arepalli, S.; Dillman, A.; Rafferty, I.P.; Troncoso, J.; et al. Abundant quantitative trait loci exist for DNA methylation and gene expression in human brain. *PLoS Genet.* **2010**, *6*, e1000952. [[CrossRef](#)] [[PubMed](#)]
27. Shen, J.; Wang, S.; Zhang, Y.J.; Kappil, M.; Wu, H.C.; Kibriya, M.G.; Wang, Q.; Jasmine, F.; Ahsan, H.; Lee, P.H.; et al. Genome-wide DNA methylation profiles in hepatocellular carcinoma. *Hepatology* **2012**, *55*, 1799–1808. [[CrossRef](#)]
28. Gopalan, S.; Carja, O.; Fagny, M.; Patin, E.; Myrick, J.W.; McEwen, L.M.; Mah, S.M.; Kobor, M.S.; Froment, A.; Feldman, M.W.; et al. Trends in DNA methylation with age replicate across diverse human populations. *Genetics* **2017**, *206*, 1659–1674. [[CrossRef](#)]
29. Liu, J.; Morgan, M.; Hutchison, K.; Calhoun, V.D. A study of the influence of sex on genome wide methylation. *PLoS ONE* **2010**, *5*, e10028. [[CrossRef](#)]
30. Zhuang, J.; Jones, A.; Lee, S.H.; Ng, E.; Fiegl, H.; Zikan, M.; Cibula, D.; Sargent, A.; Salvesen, H.B.; Jacobs, I.J.; et al. The dynamics and prognostic potential of DNA methylation changes at stem cell gene loci in women's cancer. *PLoS Genet.* **2012**, *8*, e1002517. [[CrossRef](#)]
31. Hair, B.Y.; Xu, Z.; Kirk, E.L.; Harlid, S.; Sandhu, R.; Robinson, W.R.; Wu, M.C.; Olshan, A.F.; Conway, K.; Taylor, J.A.; et al. Body mass index associated with genome-wide methylation in breast tissue. *Breast Cancer Res. Treat.* **2015**, *151*, 453–463. [[CrossRef](#)] [[PubMed](#)]

32. Qu, X.; Sandmann, T.; Frierson, H.; Fu, L.; Fuentes, E.; Walter, K.; Okrah, K.; Rumpel, C.; Moskaluk, C.; Lu, S.; et al. Integrated genomic analysis of colorectal cancer progression reveals activation of EGFR through demethylation of the EREG promoter. *Oncogene* **2016**, *35*, 6403–6415. [[CrossRef](#)] [[PubMed](#)]
33. Noreen, F.; Roosli, M.; Gaj, P.; Pietrzak, J.; Weis, S.; Urfer, P.; Regula, J.; Schar, P.; Truninger, K. Modulation of age- and cancer-associated DNA methylation change in the healthy colon by aspirin and lifestyle. *J. Natl. Cancer Inst.* **2014**, *106*. [[CrossRef](#)] [[PubMed](#)]
34. Horvath, S.; Erhart, W.; Brosch, M.; Ammerpohl, O.; von Schonfels, W.; Ahrens, M.; Heits, N.; Bell, J.T.; Tsai, P.C.; Spector, T.D.; et al. Obesity accelerates epigenetic aging of human liver. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 15538–15543. [[CrossRef](#)] [[PubMed](#)]
35. Ahrens, M.; Ammerpohl, O.; von Schonfels, W.; Kolarova, J.; Bens, S.; Itzel, T.; Teufel, A.; Herrmann, A.; Brosch, M.; Hinrichsen, H.; et al. DNA methylation analysis in nonalcoholic fatty liver disease suggests distinct disease-specific and remodeling signatures after bariatric surgery. *Cell Metab.* **2013**, *18*, 296–302. [[CrossRef](#)]
36. Kajiura, K.; Masuda, K.; Naruto, T.; Kohmoto, T.; Watabnabe, M.; Tsuboi, M.; Takizawa, H.; Kondo, K.; Tangoku, A.; Imoto, I. Frequent silencing of the candidate tumor suppressor TRIM58 by promoter methylation in early-stage lung adenocarcinoma. *Oncotarget* **2017**, *8*, 2890–2905. [[CrossRef](#)]
37. Hong, S.R.; Jung, S.E.; Lee, E.H.; Shin, K.J.; Yang, W.I.; Lee, H.Y. DNA methylation-based age prediction from saliva: High age predictability by combination of 7 CpG markers. *Forensic Sci. Int. Genet.* **2017**, *29*, 118–125. [[CrossRef](#)]
38. Timp, W.; Bravo, H.C.; McDonald, O.G.; Goggins, M.; Umbricht, C.; Zeiger, M.; Feinberg, A.P.; Irizarry, R.A. Large hypomethylated blocks as a universal defining epigenetic alteration in human solid tumors. *Genome Med.* **2014**, *6*, 61. [[CrossRef](#)]
39. Weisenberger, D.; Van Den Berg, D.; Pan, F.; Berman, B.; Laird, P.J.I. *Comprehensive DNA Methylation Analysis on the Illumina Infinium Assay Platform*; Illumina: San Diego, CA, USA, 2008.
40. Illumina. Infinium HumanMethylation450 BeadChip. Available online: [https://support.illumina.com/array/array\\_kits/infinium\\_humanmethylation450\\_beadchip\\_kit/documentation.html](https://support.illumina.com/array/array_kits/infinium_humanmethylation450_beadchip_kit/documentation.html) (accessed on 4 August 2018).
41. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
42. Benjamini, Y.; Hochberg, Y.J. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. R. Stat. Soc.* **1995**, *57*, 289–300. [[CrossRef](#)]
43. Lomax, R.G.; Hahs-Vaughn, D.L. *Statistical Concepts: A Second Course*; Routledge: New York, NY, USA, 2013.
44. Zou, H.; Hastie, T. Regularization and variable selection via the elastic net. *J. R. Stat. Soc. Ser. B (Stat. Methodol.)* **2005**, *67*, 768. [[CrossRef](#)]
45. Cortes, C.; Vapnik, V.J. Support-vector networks. *Mach. Learn.* **1995**, *20*, 273–297. [[CrossRef](#)]
46. Geisser, S. *Predictive Inference*; Routledge: London, UK, 2017.
47. Kohavi, R. A study of cross-validation and bootstrap for accuracy estimation and model selection. Proceedings of the 14th International Joint Conference on Artificial Intelligence, Montreal, QC, Canada, 20–25 August 1995; pp. 1137–1145.
48. Devijver, P.A.; Kittler, J. *Pattern Recognition: A statistical Approach*; Prentice Hall: Upper Saddle River, NJ, USA, 1982.
49. Tofallis, C. A better measure of relative prediction accuracy for model selection and model estimation. *J. Operat. Res. Soc.* **2015**, *66*, 1352–1362. [[CrossRef](#)]
50. Theil, H. *Economic Forecasts and Policy*; Cambridge University Press: Cambridge, UK, 1958.
51. Margulies, E.H.; Blanchette, M.; Haussler, D.; Green, E.D.; Progra, N.C.S. Identification and characterization of multi-species conserved sequences. *Genome Res.* **2003**, *13*, 2507–2518. [[CrossRef](#)]
52. Bindea, G.; Mlecnik, B.; Hackl, H.; Charoentong, P.; Tosolini, M.; Kirilovsky, A.; Fridman, W.H.; Pages, F.; Trajanoski, Z.; Galon, J. ClueGO: A Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics* **2009**, *25*, 1091–1093. [[CrossRef](#)]
53. Xu, C.; Qu, H.; Wang, G.; Xie, B.; Shi, Y.; Yang, Y.; Zhao, Z.; Hu, L.; Fang, X.; Yan, J.; et al. A novel strategy for forensic age prediction by DNA methylation and support vector regression model. *Sci. Rep.* **2015**, *5*, 17788. [[CrossRef](#)]

54. Jones, M.J.; Goodman, S.J.; Kobor, M.S. DNA methylation and healthy human aging. *Aging Cell* **2015**, *14*, 924–932. [[CrossRef](#)]
55. Ishimoto, Y.; Tomiyama-Miyaji, C.; Watanabe, H.; Yokoyama, H.; Ebe, K.; Tsubata, S.; Aoyagi, Y.; Abo, T. Age-dependent variation in the proportion and number of intestinal lymphocyte subsets, especially natural killer T cells, double-positive CD4+ CD8+ cells and B220+ T cells, in mice. *Immunology* **2004**, *113*, 371–377. [[CrossRef](#)]
56. Horvath, S.; Mah, V.; Lu, A.T.; Woo, J.S.; Choi, O.W.; Jasinska, A.J.; Riancho, J.A.; Tung, S.; Coles, N.S.; Braun, J.; et al. The cerebellum ages slowly according to the epigenetic clock. *Aging-Us* **2015**, *7*, 294–306. [[CrossRef](#)]
57. Sehl, M.E.; Henry, J.E.; Storniolo, A.M.; Ganz, P.A.; Horvath, S. DNA methylation age is elevated in breast tissue of healthy women. *Breast Cancer Res. Treat.* **2017**, *164*, 209–219. [[CrossRef](#)]
58. Lee, J.-H.; Park, S.-J.; Nakai, K.J. Differential landscape of non-CpG methylation in embryonic stem cells and neurons caused by DNMT3s. *Sci. Rep.* **2017**, *7*, 11295. [[CrossRef](#)] [[PubMed](#)]
59. Kim, S.; Wyckoff, J.; Morris, A.T.; Succop, A.; Avery, A.; Duncan, G.E.; Jazwinski, S.M. DNA methylation associated with healthy aging of elderly twins. *Geroscience* **2018**, *40*, 469–484. [[CrossRef](#)] [[PubMed](#)]
60. Cole, J.H.; Ritchie, S.J.; Bastin, M.E.; Valdes Hernandez, M.C.; Munoz Maniega, S.; Royle, N.; Corley, J.; Pattie, A.; Harris, S.E.; Zhang, Q.; et al. Brain age predicts mortality. *Mol. Psychiatry* **2018**, *23*, 1385–1392. [[CrossRef](#)]
61. Steiner, L.; Hopp, L.; Wirth, H.; Galle, J.; Binder, H.; Prohaska, S.J.; Rohlf, T. A global genome segmentation method for exploration of epigenetic patterns. *PLoS ONE* **2012**, *7*, e46811. [[CrossRef](#)]
62. Mamoshina, P.; Kochetov, K.; Putin, E.; Cortese, F.; Aliper, A.; Lee, W.S.; Ahn, S.M.; Uhn, L.; Skjodt, N.; Kovalchuk, O.; et al. Population Specific Biomarkers of Human Aging: A Big Data Study Using South Korean, Canadian, and Eastern European Patient Populations. *J. Gerontol. A Biol. Sci. Med. Sci.* **2018**, *73*, 1482–1490. [[CrossRef](#)] [[PubMed](#)]
63. Pyrkov, T.V.; Slipensky, K.; Barg, M.; Kondrashin, A.; Zhurov, B.; Zenin, A.; Pyatnitskiy, M.; Menshikov, L.; Markov, S.; Fedichev, P.O. Extracting biological age from biomedical data via deep learning: Too much of a good thing? *Sci. Rep.* **2018**, *8*, 5210. [[CrossRef](#)]



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