



Article

Age Prediction Using DNA Methylation Heterogeneity Metrics

Dmitry I. Karetnikov¹, Stanislav E. Romanov^{2,3} , Vladimir P. Baklaushev^{4,5,6} and Petr P. Laktionov^{2,3,*}

¹ Federal Research Center Institute of Cytology and Genetics SB RAS, 630090 Novosibirsk, Russia

² Epigenetics Laboratory, Department of Natural Sciences, Novosibirsk State University, 630090 Novosibirsk, Russia

³ Institute of Molecular and Cellular Biology, Siberian Branch of the Russian Academy of Sciences, 630090 Novosibirsk, Russia

⁴ Federal Center for Brain and Neurotechnologies, Federal Medical and Biological Agency of Russia, 117513 Moscow, Russia

⁵ Engelhardt Institute of Molecular Biology, Russian Academy of Sciences, 119991 Moscow, Russia

⁶ Department of Medical Nanobiotechnology, Medical and Biological Faculty, Pirogov Russian National Research Medical University, Ministry of Health of the Russian Federation, 117997 Moscow, Russia

* Correspondence: laktionov@mcb.nsc.ru

Abstract: Dynamic changes in genomic DNA methylation patterns govern the epigenetic developmental programs and accompany the organism's aging. Epigenetic clock (eAge) algorithms utilize DNA methylation to estimate the age and risk factors for diseases as well as analyze the impact of various interventions. High-throughput bisulfite sequencing methods, such as reduced-representation bisulfite sequencing (RRBS) or whole genome bisulfite sequencing (WGBS), provide an opportunity to identify the genomic regions of disordered or heterogeneous DNA methylation, which might be associated with cell-type heterogeneity, DNA methylation erosion, and allele-specific methylation. We systematically evaluated the applicability of five scores assessing the variability of methylation patterns by evaluating within-sample heterogeneity (WSH) to construct human blood epigenetic clock models using RRBS data. The best performance was demonstrated by the model based on a metric designed to assess DNA methylation erosion with an MAE of 3.686 years. We also trained a prediction model that uses the average methylation level over genomic regions. Although this region-based model was relatively more efficient than the WSH-based model, the latter required the analysis of just a few short genomic regions and, therefore, could be a useful tool to design a reduced epigenetic clock that is analyzed by targeted next-generation sequencing.

Keywords: DNA methylation heterogeneity; epigenetic age; bisulfite sequencing; eAge clocks



Citation: Karetnikov, D.I.; Romanov, S.E.; Baklaushev, V.P.; Laktionov, P.P. Age Prediction Using DNA Methylation Heterogeneity Metrics. *Int. J. Mol. Sci.* **2024**, *25*, 4967. <https://doi.org/10.3390/ijms25094967>

Academic Editor: David Della-Morte

Received: 30 March 2024

Revised: 27 April 2024

Accepted: 30 April 2024

Published: 2 May 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 considered a key epigenetic mark that plays a role in the regulation of gene expression, chromatin functions, genome stability, and spatial nuclear architecture [1]. Since DNA methylation dynamics accompany normal development and pathological processes, specific methylation patterns may serve as hallmarks of different cell types in their particular state [2–4]. Therefore, with the development of high-throughput sequencing and methylation microarrays, the analysis of DNA methylation is currently widely used for disease diagnostics as well as the assessment of health risks and aging [5–7]. More specifically, algorithms of epigenetic age (eAge) clocks analyze DNA methylation to predict chronological and biological age, and they serve as powerful tools for the assessment of various age-related health risks, including diseases, mental disorders, and all-cause mortality [2,8].

There are currently a few types of eAge clocks that have been developed for humans, mice, and other animals [7]. Most human eAge clocks are based on the regression modeling of microarray data that provides information about the individual methylation levels for thousands of predefined CpGs simultaneously [7]. While initial models predicted

age by several hundreds of CpGs, the detailed analysis of high-throughput methylation arrays enables researchers to restrict the list of diagnostically relevant CpGs to a few items, thereby making it possible to analyze samples with cost-effective real-time PCR, SnapShot, pyrosequencing, or targeted bisulfite sequencing (BS-seq) methods [7].

Although microarrays are powerful tools for developing methylation clocks, this technology has several drawbacks. First, widespread commercial high-throughput human microarrays are limited to predefined sets of up to 860,000 CpGs that cover around 3% of all CpG sites in the human genome [9]. Second, microarray analysis measures the methylation level of each CpG independently and is unable to extract contiguous patterns of DNA methylation as they are considered functional regulatory units [10]. These particular limitations can be overcome by using whole-genome or reduced representation bisulfite sequencing (WGBS or RRBS, respectively) methods that are based on the next-generation short-read sequencing of randomly fragmented bisulfite-converted genome DNA [11]. Specifically, WGBS covers around 90% of CpGs, and RRBS enables the analysis of around 10–20% of CpG sites, depending on the sequencing depth [12]. The main disadvantage of RRBS and WGBS is the non-uniform CpG-site coverage across samples due to the difference in sequencing depth and library preparation. The resulting inter-sample variation can be a source of uncertainty when a model built on one dataset is applied to separate samples [13]. Microarray DNA methylation data are considered standardized in this regard. However, WGBS and RRBS make it possible to analyze contiguous methylation patterns, methylation haplotype blocks, and methylation heterogeneity. Given these advantages, these methods are now widely used to discriminate cell types, and they have proven to be beneficial for cancer research and “liquid biopsy” diagnostics using circulating plasma DNA samples [5,10,14,15]. However, WGBS and RRBS are rarely used for eAge clock development and analysis [13].

The data obtained with WGBS and RRBS, to some extent, are quite difficult to interpret because it does not simply aggregate the average methylation levels of individual CpGs across all cells like in DNA methylation microarrays; instead, they reflect sample-specific sequential methylation patterns with single-molecule resolution. However, since it is believed that reproducible aging-related dynamics in DNA methylation might be caused by changes in organ cell composition due to differentiation, stem, and somatic cell loss or by epigenetic drift [2,16–18], such high-resolution data could potentially improve eAge clock development. Nevertheless, the first step when building a model using sequencing data is to bring continuous methylation patterns from every individual DNA molecule to the array of scores. This straightforward solution, which relies on the aggregation of the average methylation levels of individual CpGs within overlapping reads in the manner of microarray data, has been used previously for constructing RRBS-based eAge models [13,19]. However, this approach completely loses information about the sample-specific diversity of methylation patterns.

Alternatively, one can use a set of numerical metrics called within-sample heterogeneity scores (WSH) that capture different sources of DNA methylation heterogeneity, such as cell-type composition, allele-specific DNA methylation, or DNA methylation erosion, for every single genome region [10,12,20,21]. However, until recently, it was not clear whether such metrics were suitable for estimating chronological age since, strictly speaking, they are not directly related to the total CpG methylation level. Nevertheless, the novel results obtained on the mouse RRBS data showed that the dynamics of methylation patterns, as reflected in the form of the epigenetic “disorder” criteria, are suitable for the construction of the region-based epigenetic clocks [22]. In the present paper, we wanted to examine the ability of different heterogeneity measures to predict chronological age using the RRBS dataset from human whole blood DNA.

Here, we proposed an algorithm for epigenetic clock design based on the analysis of DNA methylation heterogeneity patterns rather than the methylation level. To identify genome regions showing an age-correlated increase or decrease in heterogeneity, we applied the previously described WSH metrics: MHL (Methylation Haplotype Load), PDR

(Proportion of Discordant Reads), PM (Epipolymorphism), FDRP (Fraction of Discordant Read Pairs), and qFDRP (quantitative Fraction of Discordant Read Pairs) [12]. Genomic regions showing the strongest association between methylation heterogeneity and age were associated with genes related to aging or linked with CpGs that are included in different epigenetic age models. We built five epigenetic clock models based on the listed metrics, and the highest performance was demonstrated by the epigenetic clock algorithm based on the PDR metric, which is designed to detect DNA methylation erosion. We show that the performance of WSH-based epigenetic clocks is mildly inferior in accuracy to conventional RRBS-based eAge clocks that are built using average methylation in genomic windows. However, WSH-based clocks only use a few loci in the genome for age prediction. As a result, heterogeneity metrics may have an advantage when RRBS data are used to design a reduced epigenetic clock.

2. Results

2.1. Development of WSH-Based Regional Blood eAge Clock Models

We used five WSH metrics: MHL (Methylation Haplotype Load), PDR (Proportion of Discordant Reads), PM (Epipolymorphism), FDRP (Fraction of Discordant Read Pairs), and qFDRP (quantitative Fraction of Discordant Read Pairs) (see Figure 1 for an explanation) [10,12,20,21]. Different WSH scores enable the assessment of distinct aspects or the biological phenomena of methylation pattern changes. Intra-molecule score PDR might be considered as a metric of DNA methylation erosion since its elevated values, which are due to stochastic demethylation, have been linked to epigenetic instability in cancer cells [21]. Another intra-molecule score, MHL, captures the homogeneity of co-methylation patterns and serves as an additional indicator of DNA methylation erosion, as it takes on a maximum value of 1 when the region is fully methylated and it is strongly reduced when extended stretches of methylated DNA are disrupted by stochastic demethylation [10]. Inter-molecule score as PM is used to quantify cell-type heterogeneity and describes the DNA methylation patterns in four-CpG windows [20]. FDRP and qFDRP scores are some other metrics that are designed to capture cell-type heterogeneity that analyzes the concordance between the same CpGs within different reads [12]. It should be noted that the basic unit of heterogeneity measurement in the MHL, PDR, FDRP, and qFDRP is a single CpG site, while for the PM, it is four contiguous CpG sites, bringing the different number of analyzed features for each metric. For simplicity, we will hereinafter refer to the listed DNA methylation heterogeneity units as heterogeneity loci regardless of the metric type.

In the first step, we evaluated the relevance of WSH scores as an instrument for epigenetic clock design. We analyzed the sequencing data of 182 bisulfite-converted blood DNA samples from donors of different ages (19–56 years, mean age 28.6 years) [23]. We obtained WSH scores for around 2 million CpGs in the case of FDRP and qFDRP scores, well above 1 million CpGs for MHL and PDR, and around 0.7 million stretches comprising four CpGs in the case of PM (Table 1). In order to reduce the diversity of features by age-associated variants, we identified heterogeneity loci with a monotonic relationship between the score value and age within each WSH metric type. Therefore, the scores for each loci were correlated with age using Spearman's rank correlation coefficient, and the heterogeneity loci that showed a correlation less than $|0.25|$ were filtered out (Table 1). The mean score values for each metric, both positively and negatively correlated, exhibit a quadratic relationship with chronological age (Figure 2), thereby indicating that changes in global heterogeneity occur most rapidly in youth and slow down by old age. It is noteworthy that a nonlinear change in heterogeneity during development was observed in [22], wherein global methylation disorder was investigated in aging mice.

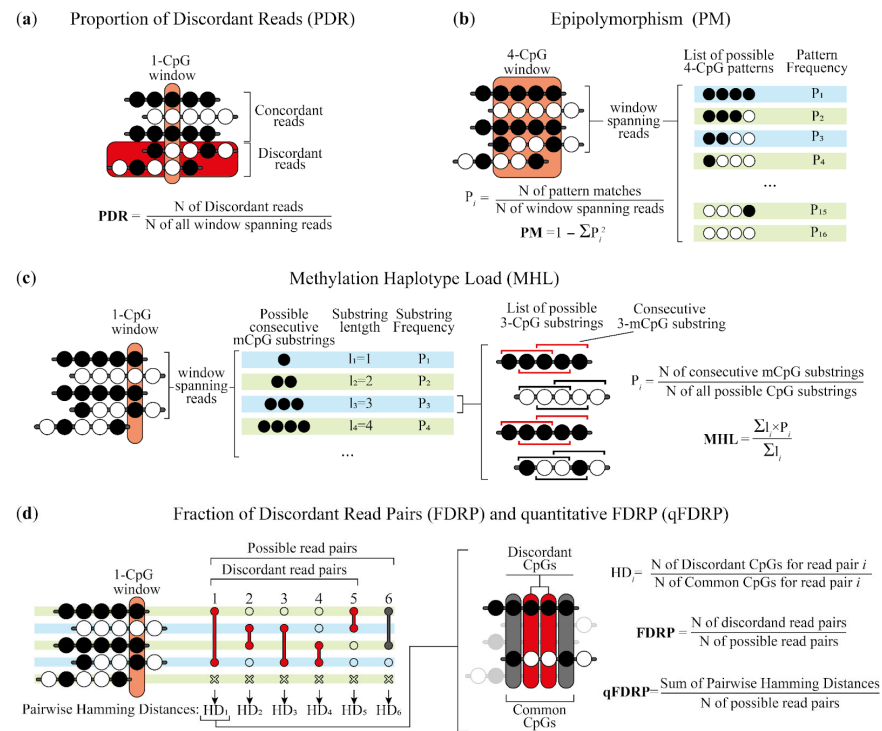


Figure 1. Schematic representation of the procedure of heterogeneity scores calculation using read sequencing data. (a) The PDR estimates the prevalence of disordered methylation patterns within CpG spanning reads in the form of the proportion of discordant reads. It classifies each read as discordant if it has both methylated and unmethylated CpGs or if it is concordant otherwise. (b) The PM reflects the amount of heterogeneity of DNA methylation for a given genomic region of 4 CpG. The metric is calculated by the frequency of all the possible 4-CpG patterns. (c) The MHL is measured by the fraction of fully methylated substrings (mCpG-substrings) of all the possible lengths throughout all the reads spanning the given CpG. It reflects how well the co-methylation haplotypes in a given region are conserved throughout the cell population. (d) The FDRP and qFDRP capture within-sample methylation heterogeneity at single CpG resolution. While FDRP calculates the fraction of discordant pairs between all the reads spanning the given CpG, qFDRP estimates the mean Hamming Distance between all the pairs (i.e., the fraction of discordant common CpGs). Note that the Hamming Distance for a concordant read pair is equal to zero.

Table 1. The diversity of heterogeneity loci for each metric in the dataset of donor blood samples. For each metric, the number of loci with a selected Spearman’s correlation is also indicated.

Heterogeneity Metrics	Number of Heterogeneity Loci	Cor ≤ −0.25	Cor ≥ 0.25	Cor ≥ 0.5
FDRP	1,918,416	2311	16,108	10
MHL	1,287,083	934	40,393	22
PDR	1,291,638	3000	25,675	48
PM	696,871	2297	11,478	20
qFDRP	1,918,416	3548	34,856	27

The resulting sets of positively and negatively correlated heterogeneity loci for each WSH score were annotated and associated with genes using ChIPseeker [24]. The total number of genes that overlapped with age-correlated heterogeneity loci is presented in Table S1. Functional annotation and GO term enrichment analysis revealed that genes associated with positively age-correlated heterogeneity loci were enriched in plenty of biological processes, the most significant being body systems development (GO:0048731, GO:0007275, GO:0048856, GO:0032502, GO:0032501), neural tissue development and differ-

entiation (GO:0007399, GO:0022008, GO:0048699, GO:0030182, GO:0048666), and cellular differentiation (GO:0030154) (Table S2). Genes associated with negatively age-correlated heterogeneity loci were significantly enriched in common terms related to organismal development (GO:0048856, GO:0032502, GO:0007275, GO:0007399, GO:0048731), signaling, and regulation of cellular communication (GO:0023051, GO:0010646) (Table S3).

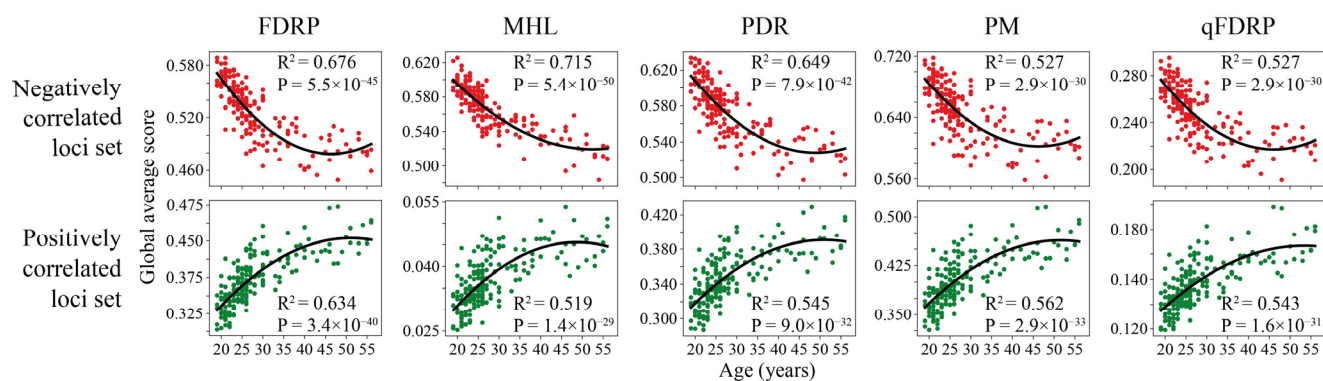


Figure 2. Age dependence of the mean heterogeneity score over positively and negatively age-correlated loci in human blood RRBS datasets. Each point corresponds to the global average score over the entire set of heterogeneity loci per single sample.

In the next step, we analyzed the heterogeneity loci, demonstrating the strongest correlation with age ($|\text{Cor}| \geq 0.5$). While the number of highly correlated heterogeneity loci varies from 10 for FDRP to 48 for PDR (Table 1), all of the loci sets are densely within a few genes, fitting into regions of no more than 300 bp in length (Tables S4–S8). Correspondingly, for different WSH scores, the number of associated genes varied between 2 and 6 genes, which often overlapped between metrics (Table 2). Interestingly, some of the listed genes are related to aging or associated with CpGs that are included in different epigenetic age models. For example, CpG sites near the genes *GRM2*, *SCGN*, and *ZIK1* are used in region-based epigenetic clocks, or they are described as age-associated CpGs [25–28]. *Lin28b* has been found to delay vasculature aging, and *ADRB1* beneficially impacts aging [29,30].

Using the corresponding highly correlated heterogeneity loci, a random forest regression model was constructed for each WSH score. Figure 3a,b shows the variances of the models on the training set. The PDR metric shows the best performance ($R^2 = 0.695$, MAE = 3.43). The PM and qFDRP metrics show relatively close performance ($R^2 = 0.595$ and 0.510 , MAE = 3.18, and 4.15, respectively), while the MHL ($R^2 = 0.436$, MAE = 4.540) and FDRP ($R^2 = 0.346$, MAE = 4.863) performed worse. Next, we evaluated the model on the test samples that were excluded from training and hyperparameter selection. The PDR metric was proven to be the most effective metric ($R^2 = 0.806$, MAE = 3.686) (Table 3, Figure 3c). Just as in the case of evaluation on the training set, the FDRP and MHL metrics showed the lowest performance on the test sample.

It is also noteworthy that the sequencing data that we used have reasonably high coverage, while the usual RRBS datasets tend to be of lower quality. To assess the applicability of the WSH scores to detect the dependencies of methylation heterogeneity change with age in RRBS-data samples, we performed a similar analysis to detect the correlations between the heterogeneity loci in RRBS-seq of mesenchymal stem cells samples (32 samples in total, age range of 0–48 years) [31]. Despite the low number of samples and single-end reads sequencing, it can be seen that there are clear associations between age and the heterogeneity scores (Figure S1). Unfortunately, due to the paucity of the samples, we were unable to build the model and assess MAE. However, the ground-age umbilical cord and placenta samples display lower levels of heterogeneity for all of the metrics used, and all of the WSH scores highly correlated with age ($R^2 > 0.8$) for both positively and negatively correlated heterogeneity loci.

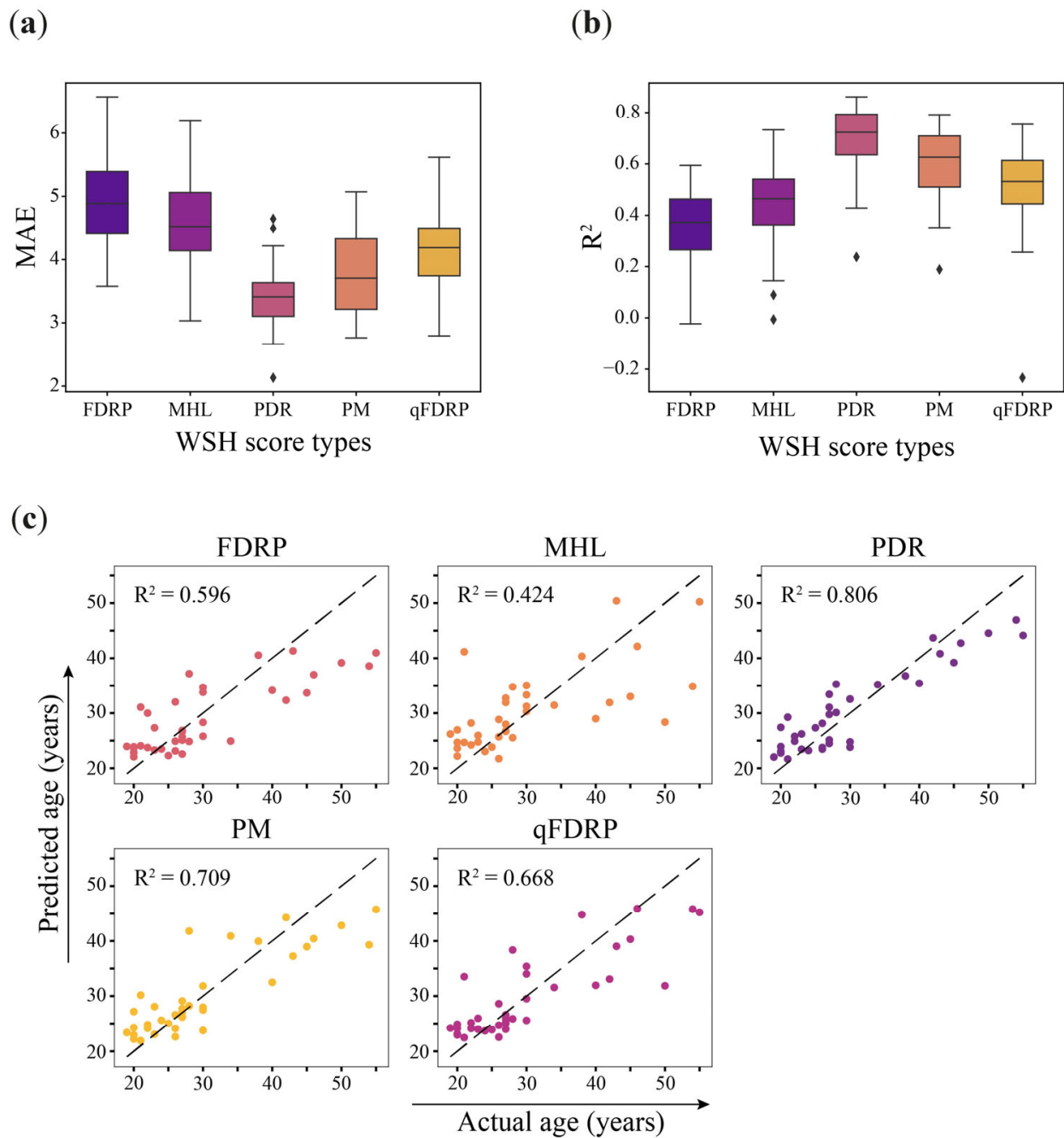


Figure 3. Performance of WSH-based RFR model. (a) Observed mean absolute error and (b) determination coefficient R^2 on the training dataset. (c) Tested performance of WSH-based eAge models.

Table 2. List of genes associated with highly correlated heterogeneity loci in all WSH scores.

Metric	Ensembl Gene ID	Gene Symbol	Gene Name	Biological Process	Gene Regions
MHL	ENSG00000164082	<i>GRM2</i>	glutamate metabotropic receptor 2	negative regulation of adenylate cyclase activity	Promoter; 5' UTR; Exon (1 of 4); Intron (1 of 5)
	ENSG00000158815	<i>FGF17</i>	fibroblast growth factor 17	positive regulation of protein phosphorylation	Exon (5 of 5)

Table 2. Cont.

Metric	Ensembl Gene ID	Gene Symbol	Gene Name	Biological Process	Gene Regions
PDR	ENSG0000043591	<i>ADRB1</i>	adrenoreceptor beta 1	positive regulation of heart rate by epinephrine-norepinephrine	Exon (1 of 1)
	ENSG0000170549	<i>IRX1</i>	Iroquois homeobox 1	negative regulation of transcription from RNA polymerase II promoter	Exon (2 of 4)
	ENSG0000122584	<i>NXPH1</i>	neurexophilin 1	NA	Intron (2 of 2)
	ENSG0000079689	<i>SCGN</i>	secretagogin, EF-hand calcium-binding protein	regulation of cytosolic calcium ion concentration	5' UTR; Exon (1 of 10)
	ENSG0000212719	<i>LINC02693</i>	long intergenic non-protein coding RNA 2693	NA	Intron (2 of 6)
	ENSG0000171649	<i>ZIK1</i>	zinc finger protein interacting with K protein	regulation of transcription from RNA polymerase II promoter	5' UTR
PM	ENSG0000150594	<i>ADRA2A</i>	adrenoreceptor alpha 2A	positive regulation of cytokine production	Exon (1 of 1)
	ENSG0000043591	<i>ADRB1</i>	adrenoreceptor beta 1	positive regulation of heart rate by epinephrine-norepinephrine	Exon (1 of 1)
	ENSG0000164082	<i>GRM2</i>	glutamate metabotropic receptor	negative regulation of adenylate cyclase activity	Promoter; Exon (1 of 4); Intron (1 of 5)
	ENSG0000212719	<i>LINC02693</i>	long intergenic non-protein coding RNA 2693	NA	Intron (2 of 6)
	ENSG0000122584	<i>NXPH1</i>	neurexophilin 1	NA	Intron (2 of 2)
FDPR	ENSG0000164082	<i>GRM2</i>	glutamate metabotropic receptor	negative regulation of adenylate cyclase activity	5' UTR; Exon (1 of 4); Intron (1 of 5)
	ENSG0000164093	<i>PITX2</i>	paired-like homeodomain	negative regulation of transcription from RNA polymerase II promoter	Exon (3 of 5)
qFDPR	ENSG0000043591	<i>ADRB1</i>	adrenoreceptor beta 1	positive regulation of heart rate by epinephrine-norepinephrine	Exon (1 of 1)
	ENSG0000164082	<i>GRM2</i>	glutamate metabotropic receptor	negative regulation of adenylate cyclase activity	Promoter; 5' UTR; Exon (1 of 4); Intron (1 of 5)
	ENSG0000187772	<i>LIN28B</i>	lin-28 homolog	miRNA catabolic process	5' UTR
	ENSG0000212719	<i>LINC02693</i>	long intergenic non-protein coding RNA 2693	NA	Intron (2 of 6)
	ENSG0000164093	<i>PITX2</i>	paired-like homeodomain	negative regulation of transcription from RNA polymerase II promoter	Exon (3 of 5)
	ENSG0000269897	<i>COMMD3-BMI1</i>	COMMD3-BMI1	sodium ion transport	Distal Intergenic

Table 3. Performance of WSH-based eAge models estimated on the test dataset.

Metrics	Test Sample	
	R ²	MAE
FDRP	0.596	4.929
MHL	0.424	5.334

Table 3. Cont.

Metrics	Test Sample	
	R ²	MAE
PDR	0.806	3.686
PM	0.709	3.969
qFDRP	0.668	4.278

2.2. Assessment of Regional Blood Epigenetic Clock Performance

To date, several approaches have been described for constructing epigenetic clock algorithms based on RRBS data, the most productive being the analysis of average methylation over genomic regions of different sizes [13,19]. To compare the efficacy of this regional approach with the heterogeneity-based one, we built an epigenetic clock model that is similar to a previously described method for mice blood RRBS eAge clocks [13]. Briefly, the average methylation frequency over genomic windows of different sizes was calculated, windows containing methylation data were deduplicated, and age-correlated windows were used for further analysis (Table S9). The hyperparameters for LASSO regression were selected on the training set and estimated with the testing set. The performance of the models depending on the window size is shown in Figure 4. It should be noted that, as the window size decreases, the model shows better R² value and MAE. The best prediction accuracy was achieved using a 100 bp sliding window with a 20 bp step size (100_20 in Figure 4 with R² = 0.885 and MAE = 2.164). Usage of windows of smaller sizes (100–1000 bp) performed better, thereby demonstrating R² in the range from 0.837 to 0.874 and MAE ranging from 2.527 to 2.266 (Figure 4, Table 4). Models based on larger genomic intervals (2000–9000 bp) showed an R² that does not exceed 0.8 and an MAE of 3 or more years. The number of regions with non-zero regression coefficients increases with decreasing window size: from 14 for the 9000 bp window to 53 for the 100 bp sliding window.

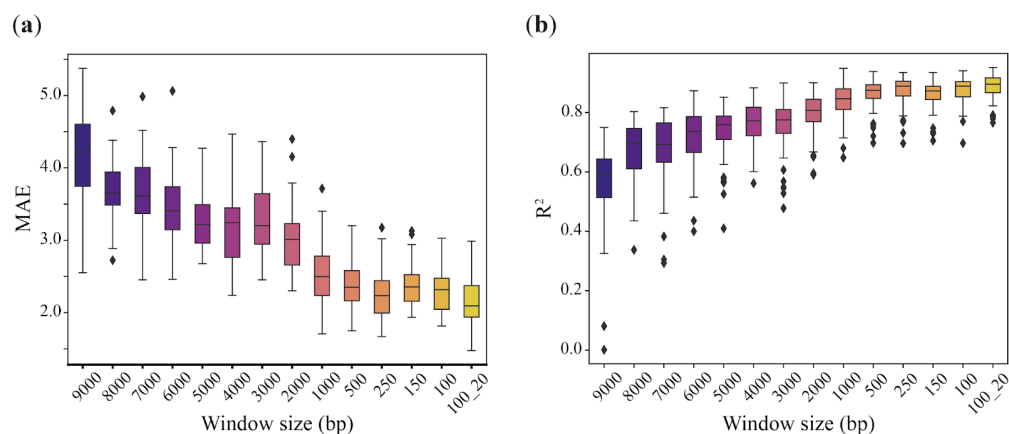


Figure 4. Training performance of regional epigenetic clocks depends on the different genomic windows that were used to calculate average DNA methylation. (a) Observed mean absolute error and (b) determination coefficient R².

Table 4. Performance of the regional epigenetic clocks models on the blood dataset.

Window Size	Cross Validation		Test Sample		Number of Regions with Non-Zero Coefficient in Regression Model
	R ²	MAE	R ²	MAE	
9000 bp	0.556	4.192	0.764	3.814	14
8000 bp	0.665	3.698	0.774	3.504	16
7000 bp	0.668	3.671	0.752	4.096	12
6000 bp	0.718	3.437	0.743	3.889	13
5000 bp	0.735	3.288	0.785	3.416	18
4000 bp	0.760	3.187	0.823	3.043	13
3000 bp	0.756	3.273	0.788	3.460	17
2000 bp	0.791	3.019	0.812	3.305	16
1000 bp	0.837	2.527	0.837	3.052	20
500 bp	0.862	2.392	0.887	2.713	28
250 bp	0.873	2.266	0.889	2.633	32
150 bp	0.858	2.379	0.884	2.750	43
100 bp	0.874	2.299	0.879	2.712	42
100 bp sliding window (20 bp step size)	0.885	2.164	0.877	2.866	53

The evaluation of the test dataset revealed the best performance of the 250 bp window model, although models built on 100–500 bp windows showed R² above 0.85 and MAE below 3 years (Table 4, Figure 5). Next, the associated genomic windows comprising 250 bp, 150 bp, and 100 bp/sliding windows models yielded a list of 52 genes, with 17 being common to all three genome segmentation approaches (Tables S10–S13). The associated genes are involved in the regulation of apoptosis, control of metabolism, cell division, and differentiation, according to the DAVID database [32].

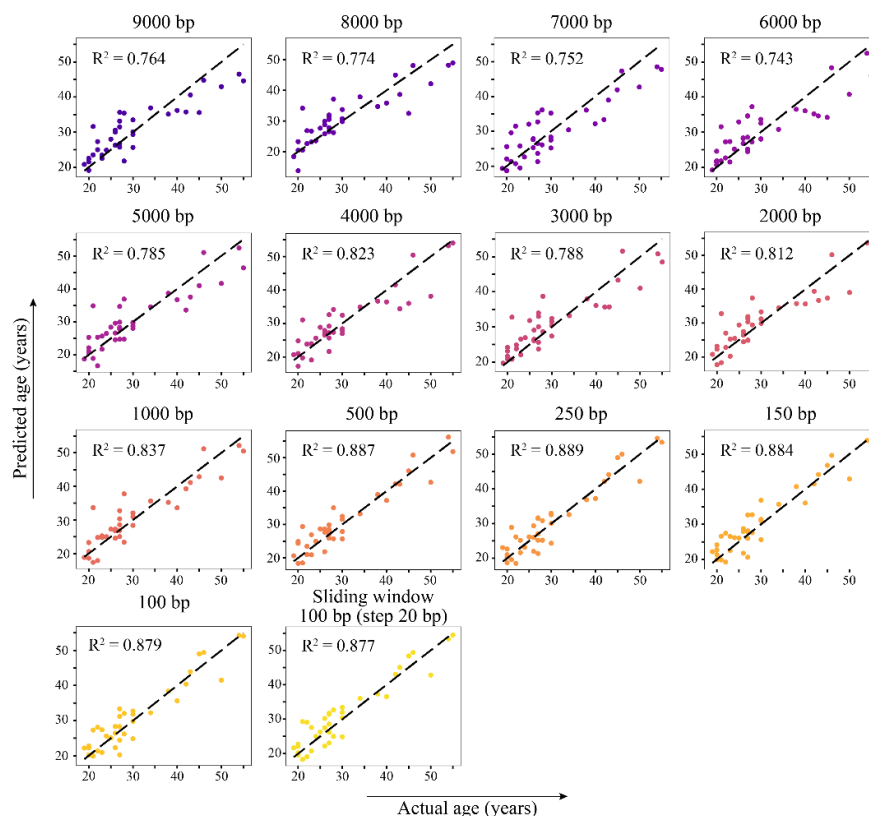


Figure 5. Performance of age prediction using regional epigenetic clocks models on blood data.

2.3. Regional Blood and WSH-Based Models for Minimized Epigenetic Clocks Design

As mentioned hereinabove, WSH-based epigenetic clock models are based on heterogeneity analysis in just a few genomic regions and, therefore, they might be considered as minimized per se. To estimate the minimum number of windows that can be used to predict age by region-based model without a loss of accuracy, we applied the Recursive Feature Elimination (RFE) method that was implemented in the scikit-learn package. For this purpose, all of the genomic windows with non-zero coefficients in the LASSO regression for 250-bp (32 regions), 150-bp (43 regions), and 100 bp sliding windows (53 regions) models were extracted from each dataset and used to retrain and test the reduced models. We consecutively reduced the number of genomic windows used in each model and tested the accuracy of the resulting eAge models (Figure 6a, Tables S14–S16). We obtained a similar performance of the 250 bp model that was reduced from 32 to 13 windows ($R^2 = 0.889$, MAE = 2.633 and $R^2 = 0.896$, MAE = 2.618, respectively). The model based on a 150 bp window was reduced up to 33 genomic regions without substantial loss in performance ($R^2 = 0.887$, MAE = 2.744). The accuracy of the 100 bp (step 20) sliding window model was conserved until the set of windows was reduced to less than 17 ($R^2 = 0.872$, MAE = 2.884) (Table S17). As long as the full 100 bp sliding window model showed the best performance in the initial setup, we analyzed the genes that localized closely to the regions included in the reduced model (Table 5). A few of them were associated with age-associated differentially methylated CpG positions in the blood (*PDCD1LG2*, *NRG2*, *C1orf132*) and with the CpGs included in other epigenetic age estimators (*C1orf132*), while the others were involved in the control of apoptosis, proliferation, and metabolism [33–38].

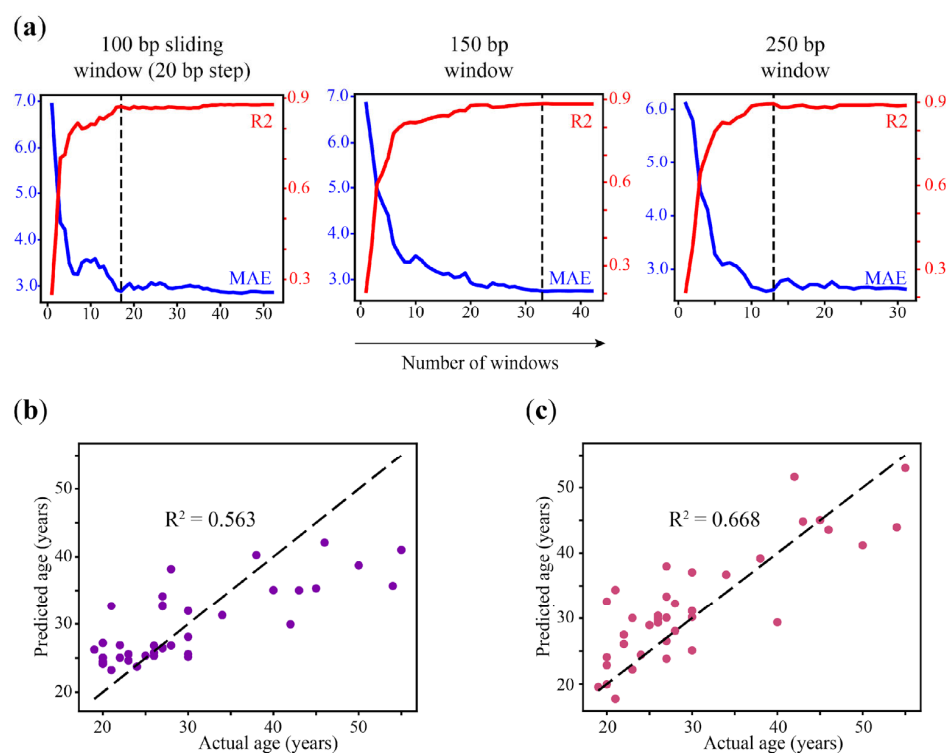


Figure 6. (a) Performance of reduced regional eAge models built. Each diagram represents the dependence of R^2 (red) and MAE (blue) from a number of windows (plotted on the X-axis) selected for model evaluation on test data. The minimum number of loci required to build a predictive model without compromising accuracy is indicated by the vertical dashed line. (b,c) Performance of the age prediction models that were designed using the reciprocal strategy of feature selection. (b) The performance of PDR-based clocks built on a set of heterogeneity loci within 100 bp (step 20 bp) sliding windows with age-correlated average methylation. (c) Performance of the region-based model built on a set of 100 bp (step 20 bp) sliding windows overlapping 48 highly age-correlated PDR loci.

Table 5. List of genes that were associated with minimized 100 bp/sliding window regional eAge model.

Ensembl Gene ID	Gene Symbol	Gene Name	Biological Process
ENSG00000114738	<i>MAPKAPK3</i>	MAPK-activated protein kinase 3	MAPK cascade
ENSG00000197142	<i>ACSL5</i>	acyl-CoA synthetase long-chain family member 5	long-chain fatty acid metabolic process
ENSG00000164082	<i>GRM2</i>	glutamate metabotropic receptor 2	negative regulation of adenylate cyclase activity
ENSG00000148734	<i>NRFFR1</i>	neuropeptide FF receptor 1	G-protein coupled receptor signaling pathway
ENSG00000115594	<i>IL1R1</i>	interleukin 1 receptor type 1	inflammatory response
ENSG00000142235	<i>LMTK3</i>	lemur tyrosine kinase 3	protein phosphorylation
ENSG00000158458	<i>NRG2</i>	neuregulin 2	signal transduction
ENSG00000103222	<i>NISCH</i>	nischarin	apoptotic process
ENSG00000197646	<i>PDCD1LG2</i>	programmed cell death 1 ligand 2	adaptive immune response
ENSG00000106772	<i>PRUNE2</i>	prune homolog 2 with BCH domain	apoptotic process
ENSG00000163239	<i>TDRD10</i>	Tudor domain containing 10	P-granule organization
ENSG00000165626	<i>BEND7</i>	BEN Domain-Containing Protein 7	
ENSG00000203709	<i>MIR29B2CHG/C1orf132</i>	MIR29B2 And MIR29C Host Gene	gene silencing by miRNA
ENSG00000236333	<i>TRHDE-AS1</i>	TRHDE antisense RNA 1	
ENSG00000166135	<i>HIF1AN</i>	hypoxia-inducible factor 1 subunit alpha inhibitor	peptidyl-histidine hydroxylation
ENSG00000180720	<i>CHRM4</i>	cholinergic receptor muscarinic 4	carbohydrate metabolic process
ENSG00000164197	<i>RNF180</i>	ring finger protein 180	protein polyubiquitination

Since the heterogeneity metrics are not directly related to the methylation level, but it does provide complementary information about the methylation pattern at loci, we wanted to test whether age prediction could be improved by combining WSH scores and average methylation by the reciprocal filtering of loci. Therefore, we first modeled heterogeneity-based age prediction using the loci from previously selected 100 bp (step 20 bp) age-correlated sliding windows (Figure 6b, Table 4). Applying heterogeneity metrics only impaired the prediction accuracy despite the fact that filtering by average methylation correlation performed best in a sliding window. We also failed to improve age prediction by generating a 100 bp step 20 sliding window region-based model on regions overlapping 48 previously selected heterogeneity loci with age-correlated PDR metrics, which demonstrated the best performance across WSH scores in the context of age prediction (Figure 6c).

Therefore, the combined approach did not improve the performance of the original models, thereby suggesting that changes in the average methylation level and DNA methylation heterogeneity with aging are not interchangeable in terms of predicting age and might detect different aspects of DNA methylation dynamics. At the same time, it should be noted that the model based on PM and qFDRP metrics shows the best performance in regions where an age-dependent methylation pattern change is observed (Table 6). Since these metrics are designed to capture DNA methylation disorder related to cell-type heterogeneity, this may imply that similar biological causes, at least to some extent, might underlie the heterogeneity changes detected by regional blood epigenetic clocks.

Table 6. The performance of WSH-based models built on genomic windows comprising 100 bp/sliding window regional eAge model.

Metrics	Test Sample	
	R ²	MAE
FDRP	0.635	4.662
MHL	0.648	4.444
PDR	0.563	4.993
PM	0.697	4.293
qFDRP	0.722	3.962

3. Discussion

The analysis of RRBS and WGBS data for epigenetic age estimation is associated with a number of technical difficulties, primarily due to the uneven coverage of the genome and the uneven representation of CpG information in different datasets. To date, several models of region-based epigenetic clocks based on the analysis of RRBS data have been proposed [13,19]. For mice, two models of epigenetic blood clocks might be considered the most effective [13]. The first model (regional-based blood clocks, RegBCs) is based on averaged methylation levels from individual CpGs within fixed-size regions that are used for LASSO regression against chronological age [13]. The second one is based on the detection of CpG clusters by the DBSCAN algorithm followed by modeling by LASSO regression (DRegBCs) [13]. Such models outperform the existing analogs based on the analysis of methylation levels of individual CpGs; for example, RegBCs showed a higher correlation with age and lower error ($R^2 = 0.91$, MAE = 3.38 months) compared to the previously described blood-specific Pekovich clock ($R^2 = 0.86$, MAE = 3.52 months) [13,39].

In our work, an approach similar to the RegBC algorithm was used as a reference, and it demonstrated high accuracy and correlation on human blood samples, achieving R^2 0.885 and MAE: 2.164, which is comparable to the most precise blood eAge models based on DNA methylation microarray data with RMSE from 2.04 years [7]. It is noteworthy that RRBS is rarely used for human epigenetic age analysis, e.g., an algorithm called intersection clocks is based on the analysis of individual CpGs, overlapped between the training and testing datasets, which maximizes the use of informative CpG sites but requires training new epigenetic clocks for each dataset [19].

The motivation for our work was to assess changes in the DNA methylation heterogeneity during aging and the applicability of such a marker for estimating age and constructing epigenetic clocks. Indeed, as mentioned above, reproducible changes in the methylation patterns are observed during aging, which might be associated with changes in the cellular composition of organs and tissues, changes in the functional state of cells, and dysfunction of DNA methylation maintenance systems [2,16–18,21,40]. Despite the lack of consensus on the physiological causes of aging that the epigenetic clock detects, a plethora of works evaluate stochastic changes in the methylation pattern, which is also referred to as epigenetic drift in the context of aging [18,41,42]. Therefore, metrics assessing DNA methylation disorder and heterogeneity, such as regional disorder and regional entropy, have been successfully used to predict age and assess the influence of common lifespan manipulation, development, and cellular dedifferentiation on epigenetic age estimates in mice [22].

Using five different WSH scores, we constructed WSH-based eAge clocks models. Interestingly, the best performance was shown by epigenetic clocks based on the PDR score, which is designed to capture methylation erosion rather than cellular heterogeneity. This finding, to some extent, echoes the observation that variability in blood cellular composition is unable to fully account for the detected methylation drift [41]. However, all heterogeneity metrics revealed disordered regions associated with genes that are related to aging or associated with CpGs that are included in different epigenetic age models. Our results show that the epigenetic clock model based on the analysis of average methylation within selected regions showed better performance than WSH-based eAge models. However, it is important to note that the minimized region-based eAge clock model requires the analysis of at least 13 genomic regions, while WSH-eAge models are based on the analysis of a smaller number of genomic regions of comparable size. It is conceivable that the WSH-eAge modeling may be the preferred approach for the design of minimized epigenetic clocks analyzed by targeted BS-seq. This requires experimental validation, but the accuracy that we have modeled is potentially consistent with the most accurate minimized eAge blood clocks that are currently available, demonstrating the error in the range from 3.5 to 8.8 years [7]. It is important to mention that WSH-eAge modeling requires relatively high genomic coverage of the regions to allow WSH score calculation; therefore, it should be applied for deep-sequenced samples or targeted bisulfite sequencing data. Another

possible issue in the usage of WSH-eAge models might arise from uneven coverage of RRBS data in separate experiments, which could require retraining of the model based on the genomic regions that are sequenced in all datasets, e.g., in a manner similar to the described algorithm of intersection clock [19].

4. Materials and Methods

4.1. Source Data

Bisulfite sequencing data of 182 human blood samples were downloaded from the ENA database via BioProject identifier PRJNA531784 [23]. In addition, a dataset containing 32 RRBS samples from mesenchymal stem cells was downloaded via BioProject identifier PRJNA349025 [31].

4.2. Data Processing and WSH Scores Calculation

Each sample was treated as follows. In the first step, adapter removal and quality control of short paired-end reads were performed using TrimGalore! v0.6.10 with `-rrbs` option [43]. The processed reads were then aligned on a GRCh38 (hg38) reference genome assembly using Bismark v0.24.2 with Bowtie 2 v2.5.2 [44]. The resulting BAM alignments were sorted by coordinate using SAMtools v1.6 [45]. The sorted BAM files were then used for the calculation of methylation heterogeneity metrics using MethEor v0.1.8 with the default parameters [46]. Overall, five metrics of heterogeneity (PM, FDRP, MHL, PDR, qFDRP) were calculated for each sample independently. The coverage files produced by Bismark, which provide information about the level of methylation for each CpG in a sample, were used for the calculation of average CpG methylation.

4.3. Heterogeneity Loci Processing and Annotation

To identify the age-related heterogeneity loci within the blood samples, Spearman's rank correlation between the given heterogeneity score and donor age, along with the FDR-adjusted p -value, was calculated for each loci using a custom script. Next, within each score type, the loci were divided into three subsets: (1) loci with a negative correlation ($\text{cor} \leq -0.25$, $\text{P}_{\text{adj}} < 0.05$), (2) loci with a positive correlation ($\text{cor} \geq 0.25$, $\text{P}_{\text{adj}} < 0.05$), and (3) with a modulus correlation above 0.5 ($|\text{cor}| \geq 0.5$, $\text{P}_{\text{adj}} < 0.05$). Each subset of loci was then annotated using ChIPseeker (distance to TSS = 0) and gProfiler, and subset (3) was used for building a score-specific Random Forest Regression (RFR) model [24,47].

4.4. WSH-Based Epigenetic Clock Construction Using Random Forest Regression

RFR modeling was performed using the scikit-learn v1.4.0 Python package [48]. In this analysis, the dataset was divided into a training (80%) and a test (20%) set of samples. The hyperparameters were selected through five-fold cross-validation on the training samples. The best hyperparameters were used to evaluate the stability and reliability of the model through five iterations of ten-fold cross-validation. In the last step, the model was evaluated on the test set, which did not undergo the training process.

4.5. Genome Segmentation and Calculation of Average Methylation Level

For a window-based estimation of CpG methylation, all CpG positions with coverage of less than five were filtered out. Next, the proportion of methylated CpGs in selected sets of genome intervals was calculated for each sample using a custom script. For these purposes, 14 sets of intervals were produced. To obtain the coordinates for each set, the genome (excluding X, Y, and M) was divided into consecutive fragments in different ways: using equal-sized bins of 100 bp–9 kb or using a sliding window algorithm (100 bp intervals with 20 bp step). The resulting BED formatted tables with the same interval set were combined by coordinates. All of the intervals without any CpG or with undetermined average methylation in at least one sample were filtered out.

To reduce the number of analyzed interval features in average methylation datasets, the Pearson correlation coefficient was calculated between the average methylation and the

age value for each interval. Intervals were filtered if their absolute correlation value was less than 0.5 ($P_{adj} < 0.05$). For the interval set, which was produced with a sliding window algorithm, deduplication was also performed to remove of identical rows. Produced data tables were then used for the LASSO regression modeling of age.

4.6. Region-Based Epigenetic Clock Construction Using LASSO Regression

To build an age prediction model on the genomic interval-based average methylation data, LASSO regression was chosen. The model was built using the scikit-learn v1.4.0 package in Python [48]. Each dataset was divided into a training set (80%) and a test set (20%). The selection of the alpha hyperparameter was carried out through ten-fold cross-validation on the training set. Next, the best hyperparameter alpha was used to build and evaluate the model on the training dataset and for evaluation on the test set.

Supplementary Materials: The following supporting information can be downloaded at <https://www.mdpi.com/article/10.3390/ijms25094967/s1>.

Author Contributions: Conceptualization, P.P.L. and V.P.B.; methodology, P.P.L.; software, D.I.K.; validation, S.E.R.; investigation, D.I.K.; resources, P.P.L.; data curation, D.I.K.; writing—original draft preparation, P.P.L. and D.I.K.; writing—review and editing, S.E.R.; visualization, S.E.R.; supervision, V.P.B.; project administration, P.P.L.; funding acquisition, P.P.L. All authors have read and agreed to the published version of the manuscript.

Funding: The research was supported by the Russian Science Foundation (project No. 22-74-10123).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data that support the findings of this study are openly available in the European Nucleotide Archive (Accessions: PRJNA531784, PRJNA349025).

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

References

- Moore, L.D.; Le, T.; Fan, G. DNA Methylation and Its Basic Function. *Neuropsychopharmacology* **2013**, *38*, 23–38. [CrossRef]
- Horvath, S.; Raj, K. DNA Methylation-Based Biomarkers and the Epigenetic Clock Theory of Ageing. *Nat. Rev. Genet.* **2018**, *19*, 371–384. [CrossRef]
- Loyfer, N.; Magenheimer, J.; Peretz, A.; Cann, G.; Bredno, J.; Klochendler, A.; Fox-Fisher, I.; Shabi-Porat, S.; Hecht, M.; Pelet, T.; et al. A DNA Methylation Atlas of Normal Human Cell Types. *Nature* **2023**, *613*, 355–364. [CrossRef] [PubMed]
- Yousefi, P.D.; Suderman, M.; Langdon, R.; Whitehurst, O.; Davey Smith, G.; Relton, C.L. DNA Methylation-Based Predictors of Health: Applications and Statistical Considerations. *Nat. Rev. Genet.* **2022**, *23*, 369–383. [CrossRef]
- Chen, X.; Gole, J.; Gore, A.; He, Q.; Lu, M.; Min, J.; Yuan, Z.; Yang, X.; Jiang, Y.; Zhang, T.; et al. Non-Invasive Early Detection of Cancer Four Years before Conventional Diagnosis Using a Blood Test. *Nat. Commun.* **2020**, *11*, 3475. [CrossRef]
- Moran, S.; Martínez-Cardús, A.; Sayols, S.; Musulén, E.; Balañá, C.; Estival-Gonzalez, A.; Moutinho, C.; Heyn, H.; Diaz-Lagares, A.; de Moura, M.C.; et al. Epigenetic Profiling to Classify Cancer of Unknown Primary: A Multicentre, Retrospective Analysis. *Lancet Oncol.* **2016**, *17*, 1386–1395. [CrossRef]
- Simpson, D.J.; Chandra, T. Epigenetic Age Prediction. *Ageing Cell* **2021**, *20*, e13452. [CrossRef] [PubMed]
- Colitã, C.-I.; Udristoiu, I.; Ancuta, D.-L.; Hermann, D.M.; Colita, D.; Colita, E.; Glavan, D.; Popa-Wagner, A. Epigenetics of Ageing and Psychiatric Disorders. *J. Integr. Neurosci.* **2024**, *23*, 13. [CrossRef] [PubMed]
- Zhou, W.; Laird, P.W.; Shen, H. Comprehensive Characterization, Annotation and Innovative Use of Infinium DNA Methylation BeadChip Probes. *Nucleic Acids Res.* **2017**, *45*, e22. [CrossRef]
- Guo, S.; Diep, D.; Plongthongkum, N.; Fung, H.-L.; Zhang, K.; Zhang, K. Identification of Methylation Haplotype Blocks Aids in Deconvolution of Heterogeneous Tissue Samples and Tumor Tissue-of-Origin Mapping from Plasma DNA. *Nat. Genet.* **2017**, *49*, 635–642. [CrossRef]
- Doherty, R.; Couldrey, C. Exploring Genome Wide Bisulfite Sequencing for DNA Methylation Analysis in Livestock: A Technical Assessment. *Front. Genet.* **2014**, *5*, 126. [CrossRef] [PubMed]
- Scherer, M.; Nebel, A.; Franke, A.; Walter, J.; Lengauer, T.; Bock, C.; Müller, F.; List, M. Quantitative Comparison of Within-Sample Heterogeneity Scores for DNA Methylation Data. *Nucleic Acids Res.* **2020**, *48*, e46. [CrossRef]
- Simpson, D.J.; Zhao, Q.; Olova, N.N.; Dabrowski, J.; Xie, X.; Latorre-Crespo, E.; Chandra, T. Region-based Epigenetic Clock Design Improves RRBS-based Age Prediction. *Ageing Cell* **2023**, *22*, e13866. [CrossRef] [PubMed]

14. Bryzgunova, O.; Bondar, A.; Ruzankin, P.; Laktionov, P.; Tarasenko, A.; Kurilshikov, A.; Epifanov, R.; Zaripov, M.; Kabilov, M.; Laktionov, P. Locus-Specific Methylation of GSTP1, RNF219, and KIAA1539 Genes with Single Molecule Resolution in Cell-Free DNA from Healthy Donors and Prostate Tumor Patients: Application in Diagnostics. *Cancers* **2021**, *13*, 6234. [[CrossRef](#)] [[PubMed](#)]
15. Mo, S.; Dai, W.; Wang, H.; Lan, X.; Ma, C.; Su, Z.; Xiang, W.; Han, L.; Luo, W.; Zhang, L.; et al. Early Detection and Prognosis Prediction for Colorectal Cancer by Circulating Tumour DNA Methylation Haplotypes: A Multicentre Cohort Study. *EClinicalMedicine* **2023**, *55*, 101717. [[CrossRef](#)]
16. Franzen, J.; Zirkel, A.; Blake, J.; Rath, B.; Benes, V.; Papantonis, A.; Wagner, W. Senescence-associated DNA Methylation Is Stochastically Acquired in Subpopulations of Mesenchymal Stem Cells. *Aging Cell* **2017**, *16*, 183–191. [[CrossRef](#)] [[PubMed](#)]
17. Hernando-Herraez, I.; Evano, B.; Stubbs, T.; Commere, P.-H.; Jan Bonder, M.; Clark, S.; Andrews, S.; Tajbakhsh, S.; Reik, W. Ageing Affects DNA Methylation Drift and Transcriptional Cell-to-Cell Variability in Mouse Muscle Stem Cells. *Nat. Commun.* **2019**, *10*, 4361. [[CrossRef](#)]
18. Yu, M.; Hazelton, W.D.; Luebeck, G.E.; Grady, W.M. Epigenetic Aging: More Than Just a Clock When It Comes to Cancer. *Cancer Res.* **2020**, *80*, 367–374. [[CrossRef](#)] [[PubMed](#)]
19. Kerepesi, C.; Gladyshev, V.N. Intersection Clock Reveals a Rejuvenation Event during Human Embryogenesis. *Aging Cell* **2023**, *22*, e13922. [[CrossRef](#)]
20. Landan, G.; Cohen, N.M.; Mukamel, Z.; Bar, A.; Molchadsky, A.; Brosh, R.; Horn-Saban, S.; Zalcenstein, D.A.; Goldfinger, N.; Zundelovich, A.; et al. Epigenetic Polymorphism and the Stochastic Formation of Differentially Methylated Regions in Normal and Cancerous Tissues. *Nat. Genet.* **2012**, *44*, 1207–1214. [[CrossRef](#)]
21. Landau, D.A.; Clement, K.; Ziller, M.J.; Boyle, P.; Fan, J.; Gu, H.; Stevenson, K.; Sougnez, C.; Wang, L.; Li, S.; et al. Locally Disordered Methylation Forms the Basis of Intratumor Methylome Variation in Chronic Lymphocytic Leukemia. *Cancer Cell* **2014**, *26*, 813–825. [[CrossRef](#)] [[PubMed](#)]
22. Bertucci-Richter, E.M.; Shealy, E.P.; Parrott, B.B. Epigenetic Drift Underlies Epigenetic Clock Signals, but Displays Distinct Responses to Lifespan Interventions, Development, and Cellular Dedifferentiation. *Aging* **2024**, *16*, 1002–1020. [[CrossRef](#)] [[PubMed](#)]
23. Bhak, Y.; Jeong, H.; Cho, Y.S.; Jeon, S.; Cho, J.; Gim, J.-A.; Jeon, Y.; Blazyte, A.; Park, S.G.; Kim, H.-M.; et al. Depression and Suicide Risk Prediction Models Using Blood-Derived Multi-Omics Data. *Transl. Psychiatry* **2019**, *9*, 262. [[CrossRef](#)] [[PubMed](#)]
24. Yu, G.; Wang, L.-G.; He, Q.-Y. ChIPseeker: An R/Bioconductor Package for ChIP Peak Annotation, Comparison and Visualization. *Bioinformatics* **2015**, *31*, 2382–2383. [[CrossRef](#)]
25. Alghanim, H.; Antunes, J.; Silva, D.S.B.S.; Alho, C.S.; Balamurugan, K.; McCord, B. Detection and Evaluation of DNA Methylation Markers Found at SCGN and KLF14 Loci to Estimate Human Age. *Forensic. Sci. Int. Genet.* **2017**, *31*, 81–88. [[CrossRef](#)] [[PubMed](#)]
26. Naue, J.; Hoefsloot, H.C.J.; Mook, O.R.F.; Rijlaarsdam-Hoekstra, L.; van der Zwalm, M.C.H.; Henneman, P.; Kloosterman, A.D.; Verschure, P.J. Chronological Age Prediction Based on DNA Methylation: Massive Parallel Sequencing and Random Forest Regression. *Forensic Sci. Int. Genet.* **2017**, *31*, 19–28. [[CrossRef](#)]
27. Tajuddin, S.M.; Hernandez, D.G.; Chen, B.H.; Noren Hooten, N.; Mode, N.A.; Nalls, M.A.; Singleton, A.B.; Ejiogu, N.; Chitrala, K.N.; Zonderman, A.B.; et al. Novel Age-Associated DNA Methylation Changes and Epigenetic Age Acceleration in Middle-Aged African Americans and Whites. *Clin. Epigenetics* **2019**, *11*, 119. [[CrossRef](#)] [[PubMed](#)]
28. Wang, Y.; Karlsson, R.; Lampa, E.; Zhang, Q.; Hedman, Å.K.; Almgren, M.; Almqvist, C.; McRae, A.F.; Marioni, R.E.; Ingelsson, E.; et al. Epigenetic Influences on Aging: A Longitudinal Genome-Wide Methylation Study in Old Swedish Twins. *Epigenetics* **2018**, *13*, 975–987. [[CrossRef](#)] [[PubMed](#)]
29. Bian, S.; Jiang, Y.; Dai, Z.; Wu, X.; Li, B.; Wang, N.; Bian, W.; Zhong, W. Lin28b Delays Vasculature Aging by Reducing Platelet-Derived Growth Factor-Beta Resistance in Senescent Vascular Smooth Muscle Cells. *Atherosclerosis* **2023**, *364*, 29–38. [[CrossRef](#)]
30. Rosoff, D.B.; Mavromatis, L.A.; Bell, A.S.; Wagner, J.; Jung, J.; Marioni, R.E.; Davey Smith, G.; Horvath, S.; Lohoff, F.W. Multivariate Genome-Wide Analysis of Aging-Related Traits Identifies Novel Loci and New Drug Targets for Healthy Aging. *Nat. Aging* **2023**, *3*, 1020–1035. [[CrossRef](#)]
31. Sheffield, N.C.; Pierron, G.; Klughammer, J.; Datlinger, P.; Schönegger, A.; Schuster, M.; Hadler, J.; Surdez, D.; Guillemot, D.; Lapouble, E.; et al. DNA Methylation Heterogeneity Defines a Disease Spectrum in Ewing Sarcoma. *Nat. Med.* **2017**, *23*, 386–395. [[CrossRef](#)] [[PubMed](#)]
32. Sherman, B.T.; Hao, M.; Qiu, J.; Jiao, X.; Baseler, M.W.; Lane, H.C.; Imamichi, T.; Chang, W. DAVID: A Web Server for Functional Enrichment Analysis and Functional Annotation of Gene Lists (2021 Update). *Nucleic Acids Res.* **2022**, *50*, W216–W221. [[CrossRef](#)]
33. Ambroa-Conde, A.; Girón-Santamaría, L.; Mosquera-Miguel, A.; Phillips, C.; Casares de Cal, M.A.; Gómez-Tato, A.; Álvarez-Dios, J.; de la Puente, M.; Ruiz-Ramírez, J.; Lareu, M.V.; et al. Epigenetic Age Estimation in Saliva and in Buccal Cells. *Forensic Sci. Int. Genet.* **2022**, *61*, 102770. [[CrossRef](#)]
34. Jung, S.-E.; Lim, S.M.; Hong, S.R.; Lee, E.H.; Shin, K.-J.; Lee, H.Y. DNA Methylation of the ELOVL2, FHL2, KLF14, C1orf132/MIR29B2C, and TRIM59 Genes for Age Prediction from Blood, Saliva, and Buccal Swab Samples. *Forensic Sci. Int. Genet.* **2019**, *38*, 1–8. [[CrossRef](#)]
35. Marttila, S.; Kananen, L.; Häyrynen, S.; Jylhävä, J.; Nevalainen, T.; Hervonen, A.; Jylhä, M.; Nykter, M.; Hurme, M. Ageing-Associated Changes in the Human DNA Methylome: Genomic Locations and Effects on Gene Expression. *BMC Genom.* **2015**, *16*, 179. [[CrossRef](#)] [[PubMed](#)]

36. Refn, M.R.; Andersen, M.M.; Kampmann, M.-L.; Tfelt-Hansen, J.; Sørensen, E.; Larsen, M.H.; Morling, N.; Børsting, C.; Pereira, V. Longitudinal Changes and Variation in Human DNA Methylation Analysed with the Illumina MethylationEPIC BeadChip Assay and Their Implications on Forensic Age Prediction. *Sci. Rep.* **2023**, *13*, 21658. [CrossRef]
37. Tharakan, R.; Ubaida-Mohien, C.; Moore, A.Z.; Hernandez, D.; Tanaka, T.; Ferrucci, L. Blood DNA Methylation and Aging: A Cross-Sectional Analysis and Longitudinal Validation in the InCHIANTI Study. *J. Gerontol. Ser. A* **2020**, *75*, 2051–2055. [CrossRef]
38. Wei, Y.; Wang, T.; Zhang, N.; Ma, Y.; Shi, S.; Zhang, R.; Zheng, X.; Zhao, L. LncRNA TRHDE-AS1 Inhibit the Scar Fibroblasts Proliferation via MiR-181a-5p/PTEN Axis. *J. Mol. Histol.* **2021**, *52*, 419–426. [CrossRef] [PubMed]
39. Petkovich, D.A.; Podolskiy, D.I.; Lobanov, A.V.; Lee, S.-G.; Miller, R.A.; Gladyshev, V.N. Using DNA Methylation Profiling to Evaluate Biological Age and Longevity Interventions. *Cell Metab.* **2017**, *25*, 954–960.e6. [CrossRef]
40. Yang, J.-H.; Hayano, M.; Griffin, P.T.; Amorim, J.A.; Bonkowski, M.S.; Apostolides, J.K.; Salfati, E.L.; Blanchette, M.; Munding, E.M.; Bhakta, M.; et al. Loss of Epigenetic Information as a Cause of Mammalian Aging. *Cell* **2023**, *186*, 305–326.e27. [CrossRef]
41. Maegawa, S.; Lu, Y.; Tahara, T.; Lee, J.T.; Madzo, J.; Liang, S.; Jelinek, J.; Colman, R.J.; Issa, J.-P.J. Caloric Restriction Delays Age-Related Methylation Drift. *Nat. Commun.* **2017**, *8*, 539. [CrossRef] [PubMed]
42. Martin, G.M. Epigenetic Gambling and Epigenetic Drift as an Antagonistic Pleiotropic Mechanism of Aging. *Aging Cell* **2009**, *8*, 761–764. [CrossRef] [PubMed]
43. Krueger, F. TrimGalore: A Wrapper around Cutadapt and FastQC to Consistently Apply Adapter and Quality Trimming to FastQ Files, with Extra Functionality for RRBS Data. Available online: <https://github.com/FelixKrueger/TrimGalore> (accessed on 1 March 2024).
44. Krueger, F.; Andrews, S.R. Bismark: A Flexible Aligner and Methylation Caller for Bisulfite-Seq Applications. *Bioinformatics* **2011**, *27*, 1571–1572. [CrossRef] [PubMed]
45. Danecek, P.; Bonfield, J.K.; Liddle, J.; Marshall, J.; Ohan, V.; Pollard, M.O.; Whitwham, A.; Keane, T.; McCarthy, S.A.; Davies, R.M.; et al. Twelve Years of SAMtools and BCFtools. *Gigascience* **2021**, *10*, giab008. [CrossRef] [PubMed]
46. Lee, D.; Koo, B.; Yang, J.; Kim, S. Methoeor: Ultrafast DNA Methylation Heterogeneity Calculation from Bisulfite Read Alignments. *PLoS Comput. Biol.* **2023**, *19*, e1010946. [CrossRef]
47. Reimand, J.; Kull, M.; Peterson, H.; Hansen, J.; Vilo, J. G:Profiler—A Web-Based Toolset for Functional Profiling of Gene Lists from Large-Scale Experiments. *Nucleic Acids Res.* **2007**, *35*, W193–W200. [CrossRef]
48. 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.

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.