
epical

Release 2.0.0

Liguo Wang, Tao Ma, et al

Apr 25, 2024

OVERVIEW

1	Introduction	1
2	Available clocks for Human	3
3	Available clocks for Mouse (<i>Mus Musculus</i>)	5
4	Installation	7
4.1	Prerequisites	7
4.2	Python Dependencies	7
4.3	Install Epical	8
4.4	Upgrade <i>Epical</i>	8
4.5	Uninstall <i>Epical</i>	8
5	Test datasets	9
5.1	Test data 1	9
5.2	Test data 2	9
5.3	Test data 3	10
5.4	Test data 4	10
5.5	Test data 5	10
5.6	Test data 6	11
5.7	Test data 7	11
6	Horvath13	13
6.1	Usage	13
6.2	Example-1	15
6.3	Example-2	16
7	Horvath13_shrunk	19
7.1	Usage	19
7.2	Example	21
8	Horvath18	23
8.1	Usage	23
8.2	Example	25
9	Levine	27
9.1	Usage	27
9.2	Example	29
10	Hannum	31
10.1	Usage	31

10.2	Example	33
11	Zhang_EN	35
11.1	Usage	35
11.2	Example	37
12	Zhang_BLUP	39
12.1	Usage	39
12.2	Example	41
13	AltumAge	43
13.1	Usage	43
13.2	Example	45
14	Cortical	47
14.1	Usage	47
14.2	Example	49
15	MEAT	51
15.1	Usage	51
15.2	Example	53
16	EPM	55
16.1	Description	55
16.2	Usage	55
16.3	Example	57
16.4	Ouput	60
17	Lu_DNAmTL	63
17.1	Usage	63
17.2	Example	65
18	PedWu	67
18.1	Usage	67
18.2	Example	69
19	PedBE	71
19.1	Usage	71
19.2	Example	73
20	GA_Bohlin	75
20.1	Usage	75
20.2	Example	77
21	GA_Haftorn	79
21.1	Usage	79
21.2	Example	81
22	GA_Knight	83
22.1	Usage	83
22.2	Example	85
23	GA_Mayne	87
23.1	Usage	87
23.2	Example	89

24	GA_Lee_CPC	91
24.1	Usage	91
24.2	Example	93
25	GA_Lee_RPC	95
25.1	Usage	95
25.2	Example	97
26	GA_Lee_rRPC	99
26.1	Usage	99
26.2	Example	101
27	WLMT	103
27.1	Usage	103
27.2	Example	105
28	YOMT	107
28.1	Usage	107
28.2	Example	109
29	mmLiver	111
29.1	Usage	111
29.2	Example	113
30	mmBlood	115
30.1	Usage	115
30.2	Example	117
31	LICENSE	119
32	Reference	121
	Index	123

INTRODUCTION

Aging is a multifaceted, time-dependent process influenced by various factors, including genetics, lifestyle, nutrition, mental well-being, as well as social and environmental conditions. Consequently, the aging speed can significantly differ among individuals, rendering **chronological age** (i.e., the number of years a person has been alive) an inadequate indicator of a person's overall health status and predictive value for disease onset and treatment responses. In contrast, **biological (or physiological) age** employs bio-physiological measurements to more accurately gauge an individual's life clock¹².

DNA methylation-based biological age estimation has been widely used; however, a universally applicable bioinformatics tool is currently lacking. The [Epical](#) package provides a number of commands to calculate epigenetic ages from DNA methylation data generated from Illumina HumanMethylation450 BeadChip (450K), MethylationEPIC v1.0 (850K) or MethylationEPIC v2.0 array.

¹ Maltoni R, Ravaoli S, Bronte G, et al. "Chronological age or biological age: What drives the choice of adjuvant treatment in elderly breast cancer patients?" *Transl Oncol* (2022).

² Rutledge J, Oh H, Wyss-Coray T. "Measuring biological age using omics data". *Nat Rev Genet* (2022).

AVAILABLE CLOCKS FOR HUMAN



	Clock_name	Predictor CpGs	Unit	Tissue	Method
1	Horvath13	353	Year	Pan-tissue	Elastic Net
2	Horvath13_shrunk	110	Year	Pan-tissue	Elastic Net
3	Horvath18	391	Year	Skin & blood	Elastic Net
4	Levine	513	Year	Blood	Elastic Net
5	Hannum	71	Year	Blood	Elastic Net
6	Zhang_EN	514	Year	Blood, Saliva	Elastic Net
7	Zhang_BLUP	319607	Year	Blood, Saliva	BLUP
8	AltumAge	20318	Year	Pan-tissue	DNN
9	Cortical	347	Year	Brain cortex	Elastic Net
10	MEAT	200	Year	Skeletal muscle	Elastic Net
11	EPM	n/a	n/a	n/a	EM
12	Lu_DNAmtL	140	Kb	Blood	Elastic Net
13	Ped_Wu	111	Year	Blood	Elastic Net
14	PedBE	94	Year	Buccal epithelial	Elastic Net
15	GA_Bohlin	96	Day	Cord blood	LASSO
16	GA_Haftorn	176	Day	Cord blood	LASSO
17	GA_Knight	148	Week	Cord blood	Elastic Net
18	GA_Mayne	62	Week	Placental tissues	Elastic Net
19	GA_Lee_CPC	546	Week	Placental tissues	Elastic Net
20	GA_Lee_RPC	558	Week	Placental tissues	Elastic Net
21	GA_Lee_rRPC	395	Week	Placental tissues	Elastic Net

Note:

- Input data for these clocks are generated from Illumina ‘BeadChip’ methylation array.
 - The “EPM” algorithm needs user provide training data.
 - All the commands (i.e. Clock_name) are case-sensitive.
-

AVAILABLE CLOCKS FOR MOUSE (*MUS MUSCULUS*)



	Clock_name	Predictor CpGs	Unit	Tissue	Method
1	WLMT	435	Day	Pan-tissue	Elastic Net
2	YOMT	329	Day	Pan-tissue	Elastic Net
3	mmLiver	148	Day	Liver	Elastic Net
4	mmBlood	90	Day	Blood	Elastic Net

Note:

- Input data for these clocks are generated from RRBS (Reduced-Representation Bisulfite Sequencing) or WGBS (Whole Genome Bisulfite Sequencing).
 - WLMT = Whole Life Multiple Tissue. This clock was trained from mice aged 6 to 30 months.
 - YOMT = YOUNG Multiple Tissue. This clock was trained from mice aged 0 to 10 months.
-

INSTALLATION

Epical is coded in Python. It needs Python 3 (version $\geq 3.5.x$ and $\leq 3.11.x$) or a later version for installation and execution.

4.1 Prerequisites

- Python 3 ($\geq 3.5.x$, $< 3.11.x$)
- pip3
- R

4.2 Python Dependencies

- pandas
- numpy
- scipy
- sklearn
- bx-python
- matplotlib
- EpigeneticPacemaker
- TensorFlow

Note: As of Jan 10, 2024. TensorFlow does NOT support Python 3.12 and 3.13

Note: Users do NOT need to install these packages manually, as they will be automatically installed if you use pip3.

4.3 Install Epical

Create virtual environment. In this example, we used `conda` to create a virtual environment named *bioage*.

```
conda create -n bioage
```

```
Collecting package metadata (current_repodata.json): done
Solving environment: done
...
```

Activate the *bioage* virtual environment

```
conda activate bioage
```

Install *epical*

```
pip3 install epical
```

```
Collecting epical
  Downloading epical-0.0.1-py3-none-any.whl.metadata (2.4 kB)
Requirement already satisfied: numpy in /Library/Frameworks/Python.framework/Versions/3.
  → 11/lib/python3.11/site-packages (from epical) (1.26.3)
Collecting scipy (from epical)
  Downloading scipy-1.11.4-cp311-cp311-macosx_10_9_x86_64.whl.metadata (60 kB)
    60.4/60.4 kB 923.3 kB/s eta 0:00:00
Collecting scikit-learn (from epical)
  Downloading scikit_learn-1.3.2-cp311-cp311-macosx_10_9_x86_64.whl.metadata (11 kB)
Collecting bx-python (from epical)
...
```

4.4 Upgrade *Epical*

```
pip3 install epical --upgrade
```

4.5 Uninstall *Epical*

```
pip3 uninstall epical
```

TEST DATASETS

We provide seven test datasets generated from using Illumina HumanMethylation450 (450K), MethylationEPIC v1.0 (850K), MethylationEPIC v2.0 and RRBS.

5.1 Test data 1

Test data 1	
Source	Mayo Clinic
Tissue	Blood
Platform	MethylationEPIC v1.0 (850K)
# of CpG	865859
# of samples	20
Age range (year)	(49, 82)
Data file	Test1_blood_N20_EPICv1_beta.tsv.gz
MD5 (Data file)	528a5896bcaae4a77757c35c7abcd507
Meta info file	Test1_blood_N20_EPICv1_info.tsv

5.2 Test data 2

Test data 2	
Source	GEO (GSE56105)
Tissue	Blood
Platform	HumanMethylation450 (450K)
# of CpG	478313
# of samples	80
Age range (year)	(10, 60)
Data file	Test2_blood_N80_450K_beta.tsv.gz
MD5 (Data file)	95e5f8ca6ab10c015644a0ff03b4b7be
Meta info file	Test2_blood_N80_450K_info.tsv

5.3 Test data 3

Test data 3	
Source	Mayo Clinic
Tissue	Liver
Platform	HumanMethylation450 (450K)
# of CpG	485512
# of samples	32
Age range (year)	(2, 81)
Data file	Test3_liver_N32_450K_beta.tsv.gz
MD5 (Data file)	a868457bd3d50aadeb5f34d50ef82dc6
Meta info file	Test3_liver_N32_450K_info.tsv

5.4 Test data 4

Test data 4	
Source	GEO (GSE229715)
Tissue	Brain
Platform	MethylationEPIC v2.0
# of CpG	853304
# of samples	16
Age range (year)	(0.4, 77)
Data file	Test4_brain_N16_EPICv2_beta.tsv.gz
MD5 (Data file)	ec566e0cf5b5dd48ed4797e132b326dd
Meta info file	Test4_brain_N16_EPICv2_info.tsv

5.5 Test data 5

Test data 5	
Source	GEO (GSE62924)
Tissue	Cord Blood
Platform	450K
# of CpG	484667
# of samples	38
Age range (week)	(34, 41)
Data file	Test5_CordBlood_N38_450K_beta.tsv.gz
MD5 (Data file)	0cf912f55a43dbbadaab8fd2371811be
Meta info file	Test5_CordBlood_N38_450K_info.tsv

5.6 Test data 6

Test data 6	
Source	GEO (GSE44667)
Tissue	Placental chorionic villus
Platform	450K
# of CpG	430685
# of samples	40
Age range (week)	(24, 38)
Data file	Test6_Placentas_N40_450K_beta.tsv.gz
MD5 (Data file)	8cb8074657a8b8d1b11e96bd1cce4165
Meta info file	Test6_Placentas_N40_450K_info.tsv

5.7 Test data 7

Test data 7	
Source	GEO (GSE121141)
Tissue	Mouse multiple tissue.
Platform	RRBS
# of CpG	994
# of samples	81
Age range (days)	(180, 900)
Data file	Test7_mouse_N81_RRBS_beta.tsv
MD5 (Data file)	98456e98a2ffddb9e78b048e048279fe
Meta info file	Test7_mouse_N81_RRBS_info.tsv

HORVATH13

The Horvath13 command calculates the DNA methylation clock developed by Steve Horvath, as published in 2013¹.

Predictor CpGs	353
Unit	Year
Target Tissue	Pan-tissue or multiple tissue
Target Population	Adult
Method	Elastic Net regression.
Reference	Horvath, S. Genome biology (2013).

6.1 Usage

```
usage: epical Horvath13 [-h] [-o out_prefix] [-p PERCENT] [-d DELIMITER]
                        [-f {pdf,png}] [-m meta_file] [-l log_file]
                        [--impute {-1,0,1,2,3,4,5,6,7,8,9,10}] [-r ref_file]
                        [--overwrite] [--debug]
                        Input_file

positional arguments:
  Input_file            The input tabular structure file containing DNA
                        methylation data. This file must have a header row,
                        which contains the names or labels for samples.
                        The first column of this file should contain CpG IDs.
                        The remaining cells in the file should contain DNA
                        methylation beta values, represented as floating-
                        point numbers between 0 and 1. Use a TAB, comma, or any
                        other delimiter to separate the columns. Use 'NaN' or
                        'NA' to represent missing values. This file can be a
                        regular text file or compressed file (".gz", ".Z",
                        ".z", ".bz", ".bz2", ".bzip2").

options:
  -h, --help            show this help message and exit
  -o out_prefix, --output out_prefix
                        The PREFIX of output files. If no PREFIX is provided,
                        the default prefix "clock_name_out" is used. The
                        generated output files include:
```

(continues on next page)

¹ Horvath S. DNA methylation age of human tissues and cell types [published correction appears in Genome Biol. 2015;16:96]. Genome Biol. 2013;14(10):R115. doi:10.1186/gb-2013-14-10-r115

(continued from previous page)

"<PREFIX>.DNAm_age.tsv": This file contains the predicted DNAm age. "<PREFIX>.used_CpGs.tsv": This file lists the CpGs that were used to calculate the DNAm age. "<PREFIX>.missed_CpGs.txt": This file provides a list of clock CpGs that were missed or excluded from the input file. "<PREFIX>.coef.tsv": This file contains a list of clock CpGs along with their coefficients. The last column indicates whether the CpG is included in the calculation.

"<PREFIX>.plots.R": This file is an R script used to generate visualization plots.

"<PREFIX>.coef_plot.pdf": This file is the coefficient plot in either PDF or PNG format.

-p PERCENT, --percent PERCENT
The maximum allowable percentage of missing CpGs. Set to 0.2 (20%) by default, which means that if more than 20% of the clock CpGs are missing, the estimation of DNAm age cannot be performed.

-d DELIMITER, --delimiter DELIMITER
Separator (usually TAB or comma) used in the input file. If the separator is not provided, the program will automatically detect the separator.

-f {pdf,png}, --format {pdf,png}
Figure format of the output coef plot. It must be "pdf" or "png". The default is "pdf".

-m meta_file, --metadata meta_file
This file contains the meta information for each sample. This file must have a header row, which contains the names or labels for variables (such as "Sex", "Age"). If the header row includes an "Age" column, a scatter plot will be generated to display the correlation between chronological age and predicted DNAm age. The first column of this file should contain sample IDs. The default value is None, indicating that no meta information file is provided.

-l log_file, --log log_file
This file is used to save the log information. By default, if no file is specified (None), the log information will be printed to the screen.

--impute {-1,0,1,2,3,4,5,6,7,8,9,10}
The imputation method code must be one of the 12 digits including (-1, 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10). The interpretations are: -1: Remove CpGs with any missing values. 0: Fill all missing values with '0.0'. 1: Fill all missing values with '1.0'. 2: Fill the missing values with **column mean** 3: Fill the missing values with **column median** 4: Fill the missing values with **column min** 5: Fill the missing values with **column max** 6: Fill the missing values with **row mean** 7: Fill the missing values with **row median** 8: Fill the missing values with **row min** 9: Fill the missing values with **row max** 10:

(continues on next page)

(continued from previous page)

```

Fill the missing values with **external reference** If
10 is specified, an external reference file must be
provided.
-r ref_file, --ref ref_file
    The external reference file contains two columns,
    separated by either tabs or commas. The first column
    represents the probe ID, while the second column
    contains the corresponding beta values.
--overwrite
    If set, over-write existing output files.
--debug
    If set, print detailed information for debugging.

```

6.2 Example-1

Input

```
$ epical Horvath13 Test1_blood_N20_EPICv1_beta.tsv.gz -o output1
```

The beta value spreadsheet ([Test1_blood_N20_EPICv1_beta.tsv.gz](#)) is the only required input file.

```

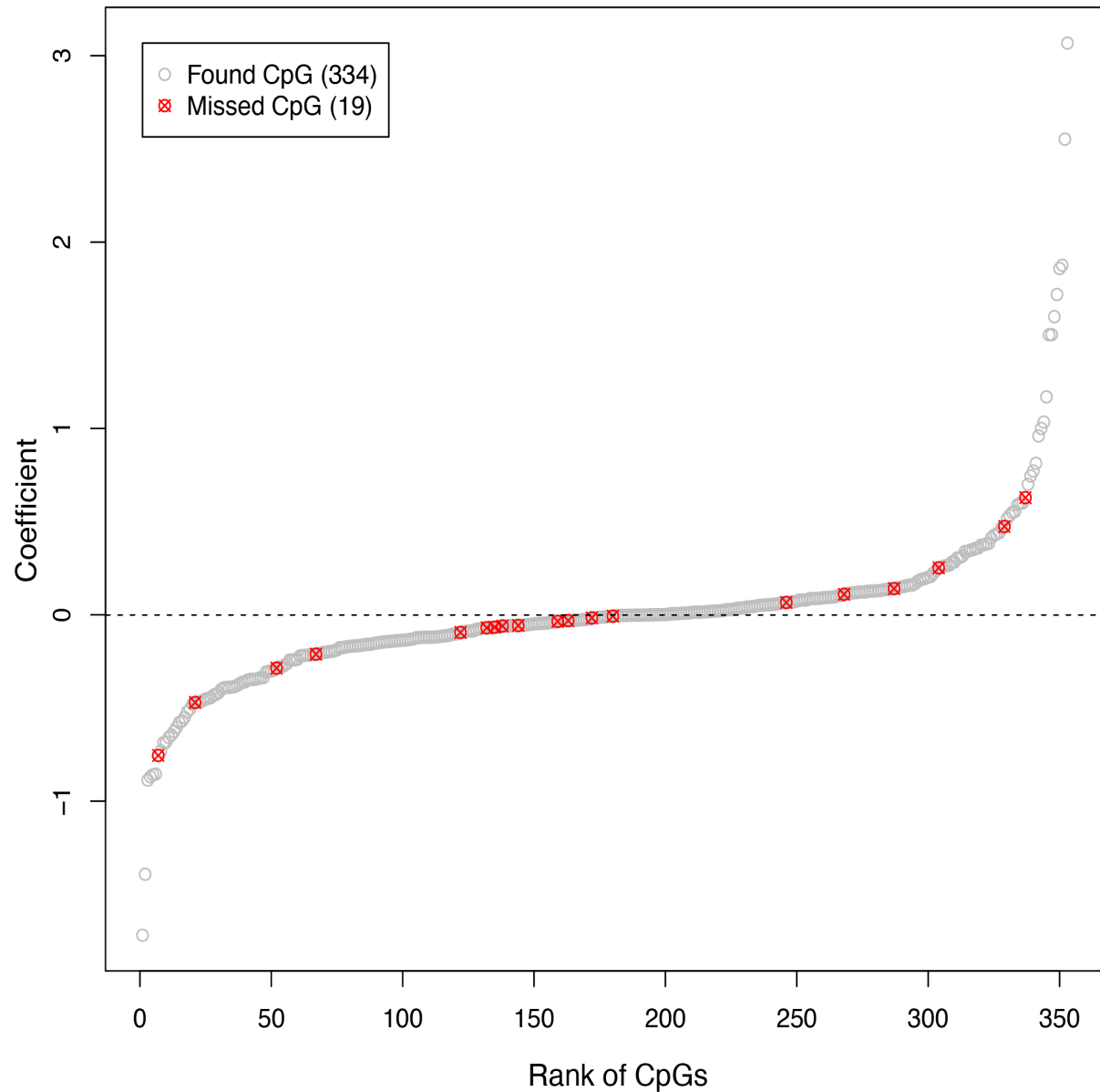
2024-01-07 09:15:01 [INFO] The prefix of output files is set to "output1".
2024-01-07 09:15:01 [INFO] Loading Horvath13 clock data ...
2024-01-07 09:15:01 [INFO] Clock's name: "Horvath13"
2024-01-07 09:15:01 [INFO] Clock was trained from: "Pan-tissue"
2024-01-07 09:15:01 [INFO] Clock's unit: "years"
2024-01-07 09:15:01 [INFO] Number of CpGs used: 353
...

```

Output

A total of 6 files are generated.

1. output1.predictorCpG_coef.tsv
 - *This file contains three columns: 1) All the predictor CpG ID; 2) Coefficients of elastic net regression; 3) Flag (True/False) indicating CpG presence in the input beta value spreadsheet.*
2. output1.predictorCpG_found.tsv
 - *Contains predictor CpGs and their beta values used for DNA methylation age calculation.*
3. output1.predictorCpG_missed.tsv
 - *Contains missed predictor CpGs from the input beta value spreadsheet.*
4. output1.DNA_m_age.tsv
 - *This file contains the sample ID (1st column) and the predicted DNA methylation age (2nd column). If a meta-information file is provided, its variables are also copied into this file (See Example-2).*
5. output1.coef_plot.pdf
 - *Ranked predictor CpG plot. All predictor CpGs were ranked by their coefficients increasingly, then missed CpGs were marked as red circles with a cross, while those used CpGs were marked as grey circles.*



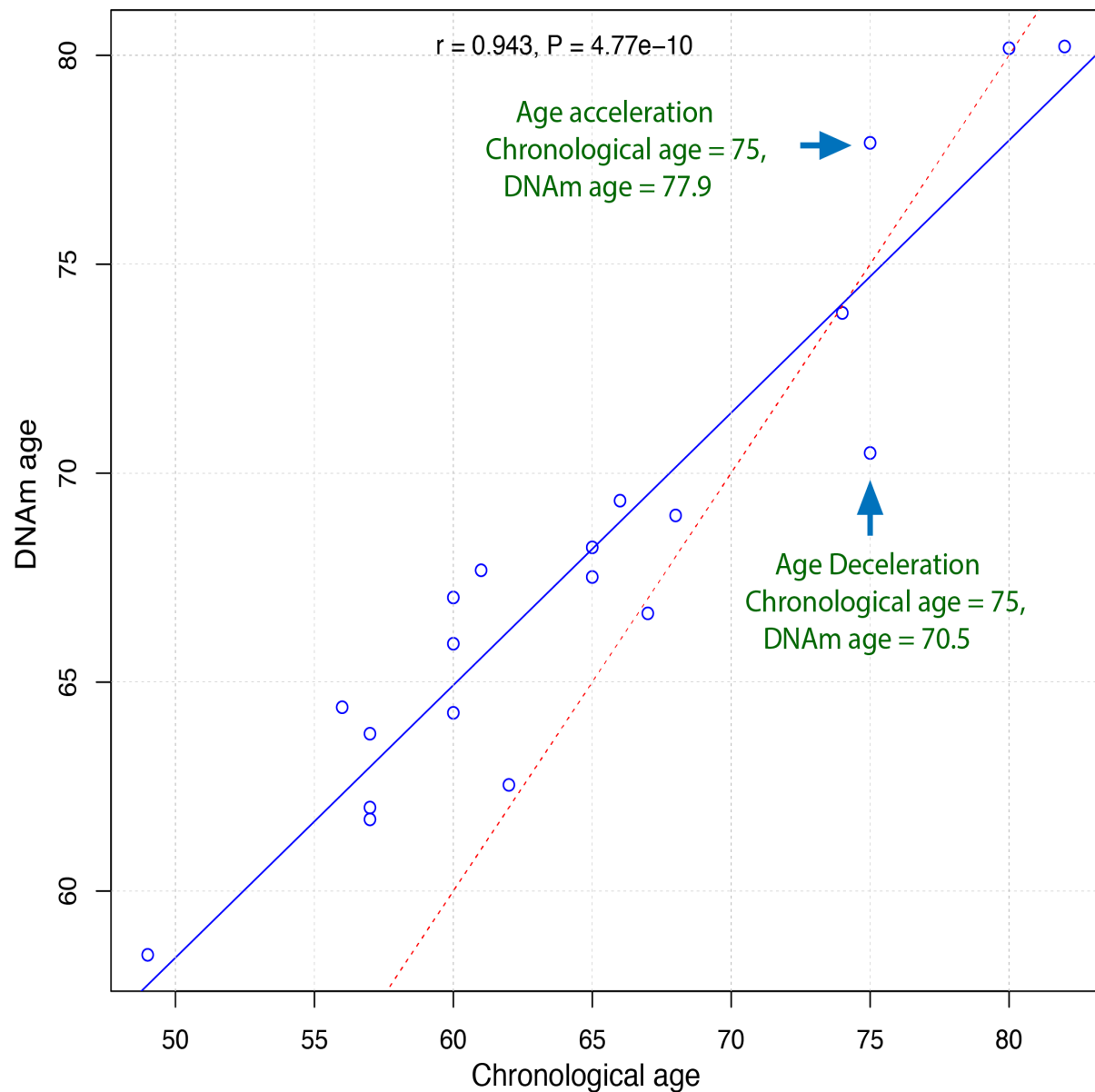
6. output1.plots.R

- R script to generate the ranked predictor CpG plot described above.

6.3 Example-2

```
$ epical Horvath13 Test1_blood_N20_EPICv1_beta.tsv.gz -m Test1_blood_N20_EPICv1_info.tsv
-o output2
```

In this case, a meta information: [Test1_blood_N20_EPICv1_info.tsv](#) file is provided. In addition to the 6 output files described above, another scatter plot will be generated showing the concordance of **chronological age** (X-axis) and **predicted age** (Y-axis).



In the plot, each dot represents a sample. The blue solid line is the linear regression line for all predicted ages. The red dashed line represents the diagonal (dots on this line have identical chronological and predicted ages). Samples above and below the red dashed line indicate age acceleration and deceleration trends, respectively.

Note:

- The sample IDs must match between the beta value spreadsheet and the meta information file.
- The meta information file must have a column named "Age".

HORVATH13_SHRUNK

The `Horvath13_shrunk` command calculates the DNA methylation using the 110 core CpGs instead of the full 353 CpGs used in the `Horvath13` command.

Predictor CpGs	110
Unit	Year
Target Tissue	Pan-tissue or multiple tissue
Target Population	Adult
Method	Elastic Net regression.
Reference	Horvath, et al., Genome biology (2013).

7.1 Usage

```
usage: epical Horvath13_shrunk [-h] [-o out_prefix] [-p PERCENT]
                               [-d DELIMITER] [-f {pdf,png}] [-m meta_file]
                               [-l log_file]
                               [--impute {-1,0,1,2,3,4,5,6,7,8,9,10}]
                               [-r ref_file] [--overwrite] [--debug]
                               Input_file

positional arguments:
  Input_file            The input tabular structure file containing DNA
                        methylation data. This file must have a header row,
                        which contains the names or labels for samples.
                        The first column of this file should contain CpG IDs.
                        The remaining cells in the file should contain DNA
                        methylation beta values, represented as floating-
                        point numbers between 0 and 1. Use a TAB, comma, or any
                        other delimiter to separate the columns. Use 'NaN' or
                        'NA' to represent missing values. This file can be a
                        regular text file or compressed file (".gz", ".Z",
                        ".z", ".bz", ".bz2", ".bzip2").

options:
  -h, --help            show this help message and exit
  -o out_prefix, --output out_prefix
                        The PREFIX of output files. If no PREFIX is provided,
                        the default prefix "clock_name_out" is used. The
```

(continues on next page)

(continued from previous page)

generated output files include:

"<PREFIX>.DNAm_age.tsv": This file contains the predicted DNAm age. "<PREFIX>.used_CpGs.tsv": This file lists the CpGs that were used to calculate the DNAm age. "<PREFIX>.missed_CpGs.txt": This file provides a list of clock CpGs that were missed or excluded from the input file. "<PREFIX>.coef.tsv": This file contains a list of clock CpGs along with their coefficients. The last column indicates whether the CpG is included in the calculation.

"<PREFIX>.plots.R": This file is an R script used to generate visualization plots.

"<PREFIX>.coef_plot.pdf": This file is the coefficient plot in either PDF or PNG format.

-p PERCENT, --percent PERCENT
The maximum allowable percentage of missing CpGs. Set to 0.2 (20%) by default, which means that if more than 20% of the clock CpGs are missing, the estimation of DNAm age cannot be performed.

-d DELIMITER, --delimiter DELIMITER
Separator (usually TAB or comma) used in the input file. If the separator is not provided, the program will automatically detect the separator.

-f {pdf,png}, --format {pdf,png}
Figure format of the output coef plot. It must be "pdf" or "png". The default is "pdf".

-m meta_file, --metadata meta_file
This file contains the meta information for each sample. This file must have a header row, which contains the names or labels for variables (such as "Sex", "Age"). If the header row includes an "Age" column, a scatter plot will be generated to display the correlation between chronological age and predicted DNAm age. The first column of this file should contain sample IDs. The default value is None, indicating that no meta information file is provided.

-l log_file, --log log_file
This file is used to save the log information. By default, if no file is specified (None), the log information will be printed to the screen.

--impute {-1,0,1,2,3,4,5,6,7,8,9,10}
The imputation method code must be one of the 12 digits including (-1, 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10). The interpretations are: -1: Remove CpGs with any missing values. 0: Fill all missing values with '0.0'. 1: Fill all missing values with '1.0'. 2: Fill the missing values with **column mean** 3: Fill the missing values with **column median** 4: Fill the missing values with **column min** 5: Fill the missing values with **column max** 6: Fill the missing values with **row mean** 7: Fill the missing values with **row median** 8: Fill the missing values with **row

(continues on next page)

(continued from previous page)

```

min** 9: Fill the missing values with **row max** 10:
Fill the missing values with **external reference** If
10 is specified, an external reference file must be
provided.
-r ref_file, --ref ref_file
    The external reference file contains two columns,
    separated by either tabs or commas. The first column
    represents the probe ID, while the second column
    contains the corresponding beta values.
--overwrite
    If set, over-write existing output files.
--debug
    If set, print detailed information for debugging.

```

7.2 Example

```
$ epical Horvath13_shrunk Test1_blood_N20_EPICv1_beta.tsv.gz -m Test1_blood_N20_EPICv1_info.
tsv -o horvath13_shrunk
```

```

2024-01-08 12:25:00 [INFO] The prefix of output files is set to "horvath13_shrunk".
2024-01-08 12:25:00 [INFO] Loading Horvath13_shrunk clock data ...
2024-01-08 12:25:00 [INFO] Clock's name: "Horvath13_shrunk"
2024-01-08 12:25:00 [INFO] Clock was trained from: "Pan-tissue"
2024-01-08 12:25:00 [INFO] Clock's unit: "years"
2024-01-08 12:25:00 [INFO] Number of CpGs used: 110
...

```

This will generate the same type of output files as the Horvath13 command.

HORVATH18

Predictor CpGs	391
Unit	Year
Target Tissue	Multiple tissue (skin & blood)
Target Population	Adult
Method	Elastic Net regression.
Reference	Horvath, et al., Aging (2018).
Alias	skin & blood clock

8.1 Usage

```
usage: epical Horvath18      [-h] [-o out_prefix] [-p PERCENT]
                             [-d DELIMITER] [-f {pdf,png}] [-m meta_file]
                             [-l log_file]
                             [--impute {-1,0,1,2,3,4,5,6,7,8,9,10}]
                             [-r ref_file] [--overwrite] [--debug]
                             Input_file

positional arguments:
  Input_file            The input tabular structure file containing DNA
                        methylation data. This file must have a header row,
                        which contains the names or labels for samples. The
                        first column of this file should contain CpG IDs. The
                        remaining cells in the file should contain DNA
                        methylation beta values, represented as floating-
                        point numbers between 0 and 1. Use a TAB, comma, or any
                        other delimiter to separate the columns. Use 'NaN' or
                        'NA' to represent missing values. This file can be a
                        regular text file or compressed file (".gz", ".Z",
                        ".z", ".bz", ".bz2", ".bzip2").

options:
  -h, --help            show this help message and exit
  -o out_prefix, --output out_prefix
                        The PREFIX of output files. If no PREFIX is provided,
                        the default prefix "clock_name_out" is used. The
                        generated output files include:
                        "<PREFIX>.DNAm_age.tsv": This file contains the
```

(continues on next page)

(continued from previous page)

predicted DNAm age. "<PREFIX>.used_CpGs.tsv": This file lists the CpGs that were used to calculate the DNAm age. "<PREFIX>.missed_CpGs.txt": This file provides a list of clock CpGs that were missed or excluded from the input file. "<PREFIX>.coef.tsv": This file contains a list of clock CpGs along with their coefficients. The last column indicates whether the CpG is included in the calculation.

"<PREFIX>.plots.R": This file is an R script used to generate visualization plots.

"<PREFIX>.coef_plot.pdf": This file is the coefficient plot in either PDF or PNG format.

-p PERCENT, --percent PERCENT
The maximum allowable percentage of missing CpGs. Set to 0.2 (20%) by default, which means that if more than 20% of the clock CpGs are missing, the estimation of DNAm age cannot be performed.

-d DELIMITER, --delimiter DELIMITER
Separator (usually TAB or comma) used in the input file. If the separator is not provided, the program will automatically detect the separator.

-f {pdf,png}, --format {pdf,png}
Figure format of the output coef plot. It must be "pdf" or "png". The default is "pdf".

-m meta_file, --metadata meta_file
This file contains the meta information for each sample. This file must have a header row, which contains the names or labels for variables (such as "Sex", "Age"). If the header row includes an "Age" column, a scatter plot will be generated to display the correlation between chronological age and predicted DNAm age. The first column of this file should contain sample IDs. The default value is None, indicating that no meta information file is provided.

-l log_file, --log log_file
This file is used to save the log information. By default, if no file is specified (None), the log information will be printed to the screen.

--impute {-1,0,1,2,3,4,5,6,7,8,9,10}
The imputation method code must be one of the 12 digits including (-1, 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10). The interpretations are: -1: Remove CpGs with any missing values. 0: Fill all missing values with '0.0'. 1: Fill all missing values with '1.0'. 2: Fill the missing values with ****column mean**** 3: Fill the missing values with ****column median**** 4: Fill the missing values with ****column min**** 5: Fill the missing values with ****column max**** 6: Fill the missing values with ****row mean**** 7: Fill the missing values with ****row median**** 8: Fill the missing values with ****row min**** 9: Fill the missing values with ****row max**** 10: Fill the missing values with ****external reference**** If

(continues on next page)

(continued from previous page)

```

10 is specified, an external reference file must be
provided.
-r ref_file, --ref ref_file
    The external reference file contains two columns,
    separated by either tabs or commas. The first column
    represents the probe ID, while the second column
    contains the corresponding beta values.
--overwrite
    If set, over-write existing output files.
--debug
    If set, print detailed information for debugging.

```

8.2 Example

```
$ epical Horvath18 Test1_blood_N20_EPICv1_beta.tsv.gz -m Test1_blood_N20_EPICv1_info.tsv
-o horvath18
```

```

2024-01-08 02:22:32 [INFO] The prefix of output files is set to "horvath18".
2024-01-08 02:22:32 [INFO] Loading Horvath18 clock data ...
2024-01-08 02:22:32 [INFO] Clock's name: "Horvath18"
2024-01-08 02:22:32 [INFO] Clock was trained from: "fibroblasts, keratinocytes, ..."
2024-01-08 02:22:32 [INFO] Clock's unit: "years"
2024-01-08 02:22:32 [INFO] Number of CpGs used: 391
...

```

This will generate the same type of output files as the Horvath13 command.

Predictor CpGs	513
Unit	Year
Target Tissue	blood, multi tissues
Target Population	Adult
Method	Elastic Net regression.
Reference	Levine ME, et al., Aging (2018).
Alias	DNAm PhenoAge

9.1 Usage

```
usage: epical Levine [-h] [-o out_prefix] [-p PERCENT] [-d DELIMITER]
                    [-f {pdf,png}] [-m meta_file] [-l log_file]
                    [--impute {-1,0,1,2,3,4,5,6,7,8,9,10}] [-r ref_file]
                    [--debug] [--overwrite]
                    Input_file

positional arguments:
  Input_file            The input tabular structure file containing DNA
                        methylation data. This file must have a header row,
                        which contains the names or labels for samples. The
                        first column of this file should contain CpG IDs. The
                        remaining cells in the file should contain DNA
                        methylation beta values, represented as floating-
                        point numbers between 0 and 1. Use a TAB, comma, or any
                        other delimiter to separate the columns. Use 'NaN' or
                        'NA' to represent missing values. This file can be a
                        regular text file or compressed file (".gz", ".Z",
                        ".z", ".bz", ".bz2", ".bzip2").

options:
  -h, --help            show this help message and exit
  -o out_prefix, --output out_prefix
                        The PREFIX of output files. If no PREFIX is provided,
                        the default prefix "clock_name_out" is used. The
                        generated output files include:
                        "<PREFIX>.DNAm_age.tsv": This file contains the
                        predicted DNAm age. "<PREFIX>.used_CpGs.tsv": This
```

(continues on next page)

(continued from previous page)

file lists the CpGs that were used to calculate the DNAm age. "<PREFIX>.missed_CpGs.txt": This file provides a list of clock CpGs that were missed or excluded from the input file. "<PREFIX>.coef.tsv": This file contains a list of clock CpGs along with their coefficients. The last column indicates whether the CpG is included in the calculation.

"<PREFIX>.plots.R": This file is an R script used to generate visualization plots.

"<PREFIX>.coef_plot.pdf": This file is the coefficient plot in either PDF or PNG format.

-p PERCENT, --percent PERCENT
The maximum allowable percentage of missing CpGs. Set to 0.2 (20%) by default, which means that if more than 20% of the clock CpGs are missing, the estimation of DNAm age cannot be performed.

-d DELIMITER, --delimiter DELIMITER
Separator (usually TAB or comma) used in the input file. If the separator is not provided, the program will automatically detect the separator.

-f {pdf,png}, --format {pdf,png}
Figure format of the output coef plot. It must be "pdf" or "png". The default is "pdf".

-m meta_file, --metadata meta_file
This file contains the meta information for each sample. This file must have a header row, which contains the names or labels for variables (such as "Sex", "Age"). If the header row includes an "Age" column, a scatter plot will be generated to display the correlation between chronological age and predicted DNAm age. The first column of this file should contain sample IDs. The default value is None, indicating that no meta information file is provided.

-l log_file, --log log_file
This file is used to save the log information. By default, if no file is specified (None), the log information will be printed to the screen.

--impute {-1,0,1,2,3,4,5,6,7,8,9,10}
The imputation method code must be one of the 12 digits including (-1, 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10). The interpretations are: -1: Remove CpGs with any missing values. 0: Fill all missing values with '0.0'. 1: Fill all missing values with '1.0'. 2: Fill the missing values with **column mean** 3: Fill the missing values with **column median** 4: Fill the missing values with **column min** 5: Fill the missing values with **column max** 6: Fill the missing values with **row mean** 7: Fill the missing values with **row median** 8: Fill the missing values with **row min** 9: Fill the missing values with **row max** 10: Fill the missing values with **external reference** If 10 is specified, an external reference file must be

(continues on next page)

(continued from previous page)

```

        provided.
-r ref_file, --ref ref_file
    The external reference file contains two columns,
    separated by either tabs or commas. The first column
    represents the probe ID, while the second column
    contains the corresponding beta values.
--debug
    If set, print detailed information for debugging.
--overwrite
    If set, over-write existing output files.

```

9.2 Example

```
$ epical Levine Test1_blood_N20_EPICv1_beta.tsv.gz -m Test1_blood_N20_EPICv1_info.tsv -o
Levine
```

```

2024-01-08 08:54:41 [INFO] The prefix of output files is set to "Levine".
2024-01-08 08:54:41 [INFO] Loading Levine clock data ...
2024-01-08 08:54:41 [INFO] Clock's name: "Levine"
2024-01-08 08:54:41 [INFO] Clock was trained from: "whole blood"
2024-01-08 08:54:41 [INFO] Clock's unit: "years"
2024-01-08 08:54:41 [INFO] Number of CpGs used: 513
...

```

This will generate the same type of output files as the Horvath13 command.

Predictor CpGs	71
Unit	Year
Target Tissue	whole blood
Target Population	Adult
Method	Elastic Net regression.
Reference	Hannum G et al., Mol Cell (2013).
Alias	

10.1 Usage

```
usage: epical Hannum [-h] [-o out_prefix] [-p PERCENT] [-d DELIMITER]
                    [-f {pdf,png}] [-m meta_file] [-l log_file]
                    [--impute {-1,0,1,2,3,4,5,6,7,8,9,10}] [-r ref_file]
                    [--debug] [--overwrite]
                    Input_file

positional arguments:
  Input_file            The input tabular structure file containing DNA
                        methylation data. This file must have a header row,
                        which contains the names or labels for samples. The
                        first column of this file should contain CpG IDs. The
                        remaining cells in the file should contain DNA
                        methylation beta values, represented as floating-
                        point numbers between 0 and 1. Use a TAB, comma, or any
                        other delimiter to separate the columns. Use 'NaN' or
                        'NA' to represent missing values. This file can be a
                        regular text file or compressed file (".gz", ".Z",
                        ".z", ".bz", ".bz2", ".bzip2").

options:
  -h, --help            show this help message and exit
  -o out_prefix, --output out_prefix
                        The PREFIX of output files. If no PREFIX is provided,
                        the default prefix "clock_name_out" is used. The
                        generated output files include:
                        "<PREFIX>.DNAm_age.tsv": This file contains the
                        predicted DNAm age. "<PREFIX>.used_CpGs.tsv": This
```

(continues on next page)

(continued from previous page)

file lists the CpGs that were used to calculate the DNAm age. "<PREFIX>.missed_CpGs.txt": This file provides a list of clock CpGs that were missed or excluded from the input file. "<PREFIX>.coef.tsv": This file contains a list of clock CpGs along with their coefficients. The last column indicates whether the CpG is included in the calculation.

"<PREFIX>.plots.R": This file is an R script used to generate visualization plots.

"<PREFIX>.coef_plot.pdf": This file is the coefficient plot in either PDF or PNG format.

-p PERCENT, --percent PERCENT
The maximum allowable percentage of missing CpGs. Set to 0.2 (20%) by default, which means that if more than 20% of the clock CpGs are missing, the estimation of DNAm age cannot be performed.

-d DELIMITER, --delimiter DELIMITER
Separator (usually TAB or comma) used in the input file. If the separator is not provided, the program will automatically detect the separator.

-f {pdf,png}, --format {pdf,png}
Figure format of the output coef plot. It must be "pdf" or "png". The default is "pdf".

-m meta_file, --metadata meta_file
This file contains the meta information for each sample. This file must have a header row, which contains the names or labels for variables (such as "Sex", "Age"). If the header row includes an "Age" column, a scatter plot will be generated to display the correlation between chronological age and predicted DNAm age. The first column of this file should contain sample IDs. The default value is None, indicating that no meta information file is provided.

-l log_file, --log log_file
This file is used to save the log information. By default, if no file is specified (None), the log information will be printed to the screen.

--impute {-1,0,1,2,3,4,5,6,7,8,9,10}
The imputation method code must be one of the 12 digits including (-1, 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10). The interpretations are: -1: Remove CpGs with any missing values. 0: Fill all missing values with '0.0'. 1: Fill all missing values with '1.0'. 2: Fill the missing values with **column mean** 3: Fill the missing values with **column median** 4: Fill the missing values with **column min** 5: Fill the missing values with **column max** 6: Fill the missing values with **row mean** 7: Fill the missing values with **row median** 8: Fill the missing values with **row min** 9: Fill the missing values with **row max** 10: Fill the missing values with **external reference** If 10 is specified, an external reference file must be

(continues on next page)

(continued from previous page)

```

        provided.
-r ref_file, --ref ref_file
    The external reference file contains two columns,
    separated by either tabs or commas. The first column
    represents the probe ID, while the second column
    contains the corresponding beta values.
--debug
    If set, print detailed information for debugging.
--overwrite
    If set, over-write existing output files.

```

10.2 Example

```
$ epical Hannum Test1_blood_N20_EPICv1_beta.tsv.gz -m Test1_blood_N20_EPICv1_info.tsv -o Hannum
```

```

2024-01-08 09:01:53 [INFO] The prefix of output files is set to "Hannum".
2024-01-08 09:01:53 [INFO] Loading Hannum clock data ...
2024-01-08 09:01:53 [INFO] Clock's name: "Hannum"
2024-01-08 09:01:53 [INFO] Clock was trained from: "whole blood"
2024-01-08 09:01:53 [INFO] Clock's unit: "years"
2024-01-08 09:01:53 [INFO] Number of CpGs used: 71
...

```


Predictor CpGs	514
Unit	Year
Target Tissue	whole blood, Saliva
Target Population	Adult
Method	Elastic Net regression.
Reference	Zhang Q, et al., <i>Genome Med.</i> (2019).
Alias	

11.1 Usage

```
usage: epical Zhang_EN [-h] [-o out_prefix] [-p PERCENT] [-d DELIMITER]
                        [-f {pdf,png}] [-m meta_file] [-l log_file]
                        [--impute {-1,0,1,2,3,4,5,6,7,8,9,10}] [-r ref_file]
                        [--debug] [--overwrite]
                        Input_file

positional arguments:
  Input_file            The input tabular structure file containing DNA
                        methylation data. This file must have a header row,
                        which contains the names or labels for samples. The
                        first column of this file should contain CpG IDs. The
                        remaining cells in the file should contain DNA
                        methylation beta values, represented as floating-
                        point numbers between 0 and 1. Use a TAB, comma, or any
                        other delimiter to separate the columns. Use 'NaN' or
                        'NA' to represent missing values. This file can be a
                        regular text file or compressed file (".gz", ".Z",
                        ".z", ".bz", ".bz2", ".bzip2").

options:
  -h, --help            show this help message and exit
  -o out_prefix, --output out_prefix
                        The PREFIX of output files. If no PREFIX is provided,
                        the default prefix "clock_name_out" is used. The
                        generated output files include:
                        "<PREFIX>.DNAm_age.tsv": This file contains the
                        predicted DNAm age. "<PREFIX>.used_CpGs.tsv": This
```

(continues on next page)

(continued from previous page)

file lists the CpGs that were used to calculate the DNAm age. "<PREFIX>.missed_CpGs.txt": This file provides a list of clock CpGs that were missed or excluded from the input file. "<PREFIX>.coef.tsv": This file contains a list of clock CpGs along with their coefficients. The last column indicates whether the CpG is included in the calculation.

"<PREFIX>.plots.R": This file is an R script used to generate visualization plots.

"<PREFIX>.coef_plot.pdf": This file is the coefficient plot in either PDF or PNG format.

-p PERCENT, --percent PERCENT
The maximum allowable percentage of missing CpGs. Set to 0.2 (20%) by default, which means that if more than 20% of the clock CpGs are missing, the estimation of DNAm age cannot be performed.

-d DELIMITER, --delimiter DELIMITER
Separator (usually TAB or comma) used in the input file. If the separator is not provided, the program will automatically detect the separator.

-f {pdf,png}, --format {pdf,png}
Figure format of the output coef plot. It must be "pdf" or "png". The default is "pdf".

-m meta_file, --metadata meta_file
This file contains the meta information for each sample. This file must have a header row, which contains the names or labels for variables (such as "Sex", "Age"). If the header row includes an "Age" column, a scatter plot will be generated to display the correlation between chronological age and predicted DNAm age. The first column of this file should contain sample IDs. The default value is None, indicating that no meta information file is provided.

-l log_file, --log log_file
This file is used to save the log information. By default, if no file is specified (None), the log information will be printed to the screen.

--impute {-1,0,1,2,3,4,5,6,7,8,9,10}
The imputation method code must be one of the 12 digits including (-1, 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10). The interpretations are: -1: Remove CpGs with any missing values. 0: Fill all missing values with '0.0'. 1: Fill all missing values with '1.0'. 2: Fill the missing values with **column mean** 3: Fill the missing values with **column median** 4: Fill the missing values with **column min** 5: Fill the missing values with **column max** 6: Fill the missing values with **row mean** 7: Fill the missing values with **row median** 8: Fill the missing values with **row min** 9: Fill the missing values with **row max** 10: Fill the missing values with **external reference** If 10 is specified, an external reference file must be

(continues on next page)

(continued from previous page)

```

        provided.
-r ref_file, --ref ref_file
    The external reference file contains two columns,
    separated by either tabs or commas. The first column
    represents the probe ID, while the second column
    contains the corresponding beta values.
--debug
    If set, print detailed information for debugging.
--overwrite
    If set, over-write existing output files.

```

11.2 Example

```

$ epical Zhang_EN Test1_blood_N20_EPICv1_beta.tsv.gz -m Test1_blood_N20_EPICv1_info.tsv
-o Zhang_EN

```

```

2024-01-17 09:41:08 [INFO] The prefix of output files is set to "Zhang_EN".
2024-01-17 09:41:08 [INFO] Loading Zhang_EN clock data ...
2024-01-17 09:41:08 [INFO] Clock's name: "Zhang_EN"
2024-01-17 09:41:08 [INFO] Clock was trained from: "blood,saliva"
2024-01-17 09:41:08 [INFO] Clock's unit: "years"
2024-01-17 09:41:08 [INFO] Number of CpGs used: 514
...

```


Predictor CpGs	319607
Unit	Year
Target Tissue	whole blood, Saliva
Target Population	Adult
Method	Best Linear Unbiased Prediction (BLUP)
Reference	Zhang Q, et al., <i>Genome Med.</i> (2019).
Alias	

12.1 Usage

```
usage: epical Zhang_BLUP [-h] [-o out_prefix] [-p PERCENT] [-d DELIMITER]
                        [-f {pdf,png}] [-m meta_file] [-l log_file]
                        [--impute {-1,0,1,2,3,4,5,6,7,8,9,10}] [-r ref_file]
                        [--debug] [--overwrite]
                        Input_file

positional arguments:
  Input_file            The input tabular structure file containing DNA
                        methylation data. This file must have a header row,
                        which contains the names or labels for samples. The
                        first column of this file should contain CpG IDs. The
                        remaining cells in the file should contain DNA
                        methylation beta values, represented as floating-
                        point numbers between 0 and 1. Use a TAB, comma, or any
                        other delimiter to separate the columns. Use 'NaN' or
                        'NA' to represent missing values. This file can be a
                        regular text file or compressed file (".gz", ".Z",
                        ".z", ".bz", ".bz2", ".bzip2").

options:
  -h, --help            show this help message and exit
  -o out_prefix, --output out_prefix
                        The PREFIX of output files. If no PREFIX is provided,
                        the default prefix "clock_name_out" is used. The
                        generated output files include:
                        "<PREFIX>.DNAm_age.tsv": This file contains the
                        predicted DNAm age. "<PREFIX>.used_CpGs.tsv": This
```

(continues on next page)

(continued from previous page)

file lists the CpGs that were used to calculate the DNAm age. "<PREFIX>.missed_CpGs.txt": This file provides a list of clock CpGs that were missed or excluded from the input file. "<PREFIX>.coef.tsv": This file contains a list of clock CpGs along with their coefficients. The last column indicates whether the CpG is included in the calculation.

"<PREFIX>.plots.R": This file is an R script used to generate visualization plots.

"<PREFIX>.coef_plot.pdf": This file is the coefficient plot in either PDF or PNG format.

-p PERCENT, --percent PERCENT
The maximum allowable percentage of missing CpGs. Set to 0.2 (20%) by default, which means that if more than 20% of the clock CpGs are missing, the estimation of DNAm age cannot be performed.

-d DELIMITER, --delimiter DELIMITER
Separator (usually TAB or comma) used in the input file. If the separator is not provided, the program will automatically detect the separator.

-f {pdf,png}, --format {pdf,png}
Figure format of the output coef plot. It must be "pdf" or "png". The default is "pdf".

-m meta_file, --metadata meta_file
This file contains the meta information for each sample. This file must have a header row, which contains the names or labels for variables (such as "Sex", "Age"). If the header row includes an "Age" column, a scatter plot will be generated to display the correlation between chronological age and predicted DNAm age. The first column of this file should contain sample IDs. The default value is None, indicating that no meta information file is provided.

-l log_file, --log log_file
This file is used to save the log information. By default, if no file is specified (None), the log information will be printed to the screen.

--impute {-1,0,1,2,3,4,5,6,7,8,9,10}
The imputation method code must be one of the 12 digits including (-1, 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10). The interpretations are: -1: Remove CpGs with any missing values. 0: Fill all missing values with '0.0'. 1: Fill all missing values with '1.0'. 2: Fill the missing values with ****column mean**** 3: Fill the missing values with ****column median**** 4: Fill the missing values with ****column min**** 5: Fill the missing values with ****column max**** 6: Fill the missing values with ****row mean**** 7: Fill the missing values with ****row median**** 8: Fill the missing values with ****row min**** 9: Fill the missing values with ****row max**** 10: Fill the missing values with ****external reference**** If 10 is specified, an external reference file must be

(continues on next page)

(continued from previous page)

```

        provided.
-r ref_file, --ref ref_file
    The external reference file contains two columns,
    separated by either tabs or commas. The first column
    represents the probe ID, while the second column
    contains the corresponding beta values.
--debug
    If set, print detailed information for debugging.
--overwrite
    If set, over-write existing output files.

```

12.2 Example

```

$ epical Zhang_BLUP Test1_blood_N20_EPICv1_beta.tsv.gz -m Test1_blood_N20_EPICv1_info.tsv
-o Zhang_BLUP

```

```

2024-01-17 09:43:54 [INFO] The prefix of output files is set to "Zhang_BLUP".
2024-01-17 09:43:54 [INFO] Loading Zhang_BLUP clock data ...
2024-01-17 09:43:54 [INFO] Clock's name: "Zhang_BLUP"
2024-01-17 09:43:54 [INFO] Clock was trained from: "blood,saliva"
2024-01-17 09:43:54 [INFO] Clock's unit: "years"
2024-01-17 09:43:54 [INFO] Number of CpGs used: 319607
...

```


ALTUMAGE

Predictor CpGs	20318
Unit	Year
Target Tissue	Pan-tissue
Target Population	Adult
Method	Deep Neural Network (DNN)
Reference	LP de Lima Camillo et al., NPJ Aging. (2022).
Alias	

13.1 Usage

```
usage: epical AltumAge [-h] [-o out_prefix] [-p PERCENT] [-d DELIMITER]
                        [-f {pdf,png}] [-m meta_file] [-l log_file]
                        [--impute {-1,0,1,2,3,4,5,6,7,8,9,10}] [-r ref_file]
                        [--debug] [--overwrite]
                        Input_file

positional arguments:
  Input_file            The input tabular structure file containing DNA
                        methylation data. This file must have a header row,
                        which contains the names or labels for samples. The
                        first column of this file should contain CpG IDs. The
                        remaining cells in the file should contain DNA
                        methylation beta values, represented as floating-
                        point numbers between 0 and 1. Use a TAB, comma, or any
                        other delimiter to separate the columns. Use 'NaN' or
                        'NA' to represent missing values. This file can be a
                        regular text file or compressed file (".gz", ".Z",
                        ".z", ".bz", ".bz2", ".bzip2").

options:
  -h, --help            show this help message and exit
  -o out_prefix, --output out_prefix
                        The PREFIX of output files. If no PREFIX is provided,
                        the default prefix "clock_name_out" is used. The
                        generated output files include:
                        "<PREFIX>.DNAm_age.tsv": This file contains the
                        predicted DNAm age. "<PREFIX>.used_CpGs.tsv": This
```

(continues on next page)

(continued from previous page)

file lists the CpGs that were used to calculate the DNAm age. "<PREFIX>.missed_CpGs.txt": This file provides a list of clock CpGs that were missed or excluded from the input file. "<PREFIX>.coef.tsv": This file contains a list of clock CpGs along with their coefficients. The last column indicates whether the CpG is included in the calculation.

"<PREFIX>.plots.R": This file is an R script used to generate visualization plots.

"<PREFIX>.coef_plot.pdf": This file is the coefficient plot in either PDF or PNG format.

-p PERCENT, --percent PERCENT
The maximum allowable percentage of missing CpGs. Set to 0.2 (20%) by default, which means that if more than 20% of the clock CpGs are missing, the estimation of DNAm age cannot be performed.

-d DELIMITER, --delimiter DELIMITER
Separator (usually TAB or comma) used in the input file. If the separator is not provided, the program will automatically detect the separator.

-f {pdf,png}, --format {pdf,png}
Figure format of the output coef plot. It must be "pdf" or "png". The default is "pdf".

-m meta_file, --metadata meta_file
This file contains the meta information for each sample. This file must have a header row, which contains the names or labels for variables (such as "Sex", "Age"). If the header row includes an "Age" column, a scatter plot will be generated to display the correlation between chronological age and predicted DNAm age. The first column of this file should contain sample IDs. The default value is None, indicating that no meta information file is provided.

-l log_file, --log log_file
This file is used to save the log information. By default, if no file is specified (None), the log information will be printed to the screen.

--impute {-1,0,1,2,3,4,5,6,7,8,9,10}
The imputation method code must be one of the 12 digits including (-1, 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10). The interpretations are: -1: Remove CpGs with any missing values. 0: Fill all missing values with '0.0'. 1: Fill all missing values with '1.0'. 2: Fill the missing values with ****column mean**** 3: Fill the missing values with ****column median**** 4: Fill the missing values with ****column min**** 5: Fill the missing values with ****column max**** 6: Fill the missing values with ****row mean**** 7: Fill the missing values with ****row median**** 8: Fill the missing values with ****row min**** 9: Fill the missing values with ****row max**** 10: Fill the missing values with ****external reference**** If 10 is specified, an external reference file must be

(continues on next page)

(continued from previous page)

```

        provided.
-r ref_file, --ref ref_file
    The external reference file contains two columns,
    separated by either tabs or commas. The first column
    represents the probe ID, while the second column
    contains the corresponding beta values.
--debug
    If set, print detailed information for debugging.
--overwrite
    If set, over-write existing output files.

```

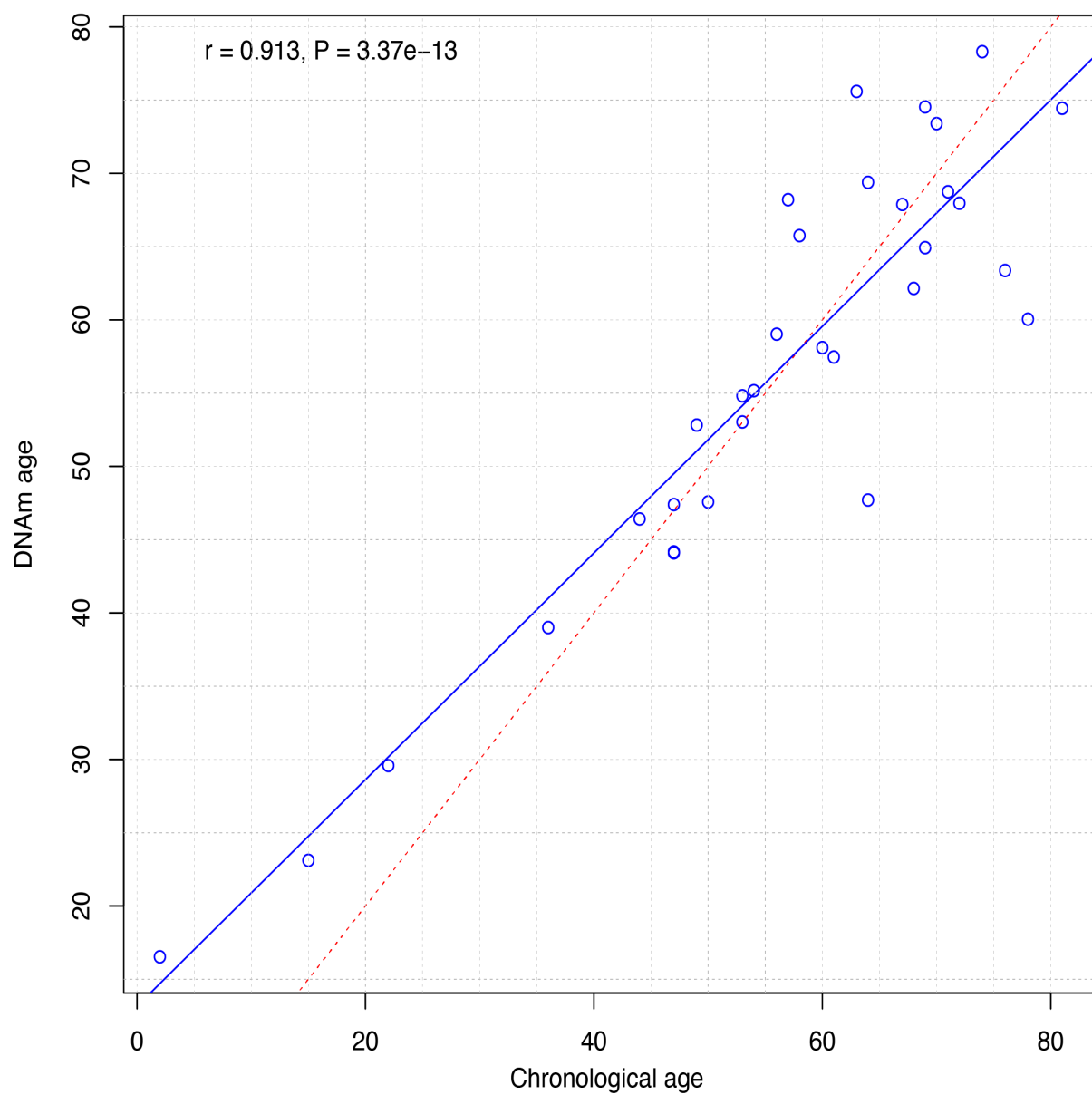
13.2 Example

```
$ epical AltumAge Test3_liver_N32_450K_beta.tsv.gz -m Test3_liver_N32_450K_info.tsv -o
AltumAge
```

```

2024-01-17 21:58:14.840540: I tensorflow/core/platform/cpu_feature_guard.cc:182] This
↳ TensorFlow binary is optimized to use available CPU instructions in performance-
↳ critical operations.
To enable the following instructions: AVX2 FMA, in other operations, rebuild TensorFlow
↳ with the appropriate compiler flags.
2024-01-17 09:58:31 [INFO] The prefix of output files is set to "AltumAge".
2024-01-17 09:58:31 [INFO] Loading AltumAge clock data ...
2024-01-17 09:58:33 [INFO] Clock's name: "AltumAge"
2024-01-17 09:58:33 [INFO] Clock was trained from: "Pan-tissue"
2024-01-17 09:58:33 [INFO] Clock's unit: "years"
2024-01-17 09:58:33 [INFO] Number of CpGs used: 20318
...

```



CORTICAL

Predictor CpGs	347
Unit	Year
Target Tissue	Brain cortex
Target Population	Adult
Method	Elastic Net regression
Reference	Shireby GL, et al, Brain. (2020).
Alias	DNAmClock-Cortical

14.1 Usage

```
usage: epical Cortical [-h] [-o out_prefix] [-p PERCENT] [-d DELIMITER]
                        [-f {pdf,png}] [-m meta_file] [-l log_file]
                        [--impute {-1,0,1,2,3,4,5,6,7,8,9,10}] [-r ref_file]
                        [--overwrite] [--debug]
                        Input_file

positional arguments:
  Input_file            The input tabular structure file containing DNA
                        methylation data. This file must have a header row,
                        which contains the names or labels for samples. The
                        first column of this file should contain CpG IDs. The
                        remaining cells in the file should contain DNA
                        methylation beta values, represented as floating-
                        point numbers between 0 and 1. Use a TAB, comma, or any
                        other delimiter to separate the columns. Use 'NaN' or
                        'NA' to represent missing values. This file can be a
                        regular text file or compressed file (".gz", ".Z",
                        ".z", ".bz", ".bz2", ".bzip2").

options:
  -h, --help            show this help message and exit
  -o out_prefix, --output out_prefix
                        The PREFIX of output files. If no PREFIX is provided,
                        the default prefix "clock_name_out" is used. The
                        generated output files include:
                        "<PREFIX>.DNAm_age.tsv": This file contains the
                        predicted DNAm age. "<PREFIX>.used_CpGs.tsv": This
```

(continues on next page)

(continued from previous page)

file lists the CpGs that were used to calculate the DNAm age. "<PREFIX>.missed_CpGs.txt": This file provides a list of clock CpGs that were missed or excluded from the input file. "<PREFIX>.coef.tsv": This file contains a list of clock CpGs along with their coefficients. The last column indicates whether the CpG is included in the calculation.

"<PREFIX>.plots.R": This file is an R script used to generate visualization plots.

"<PREFIX>.coef_plot.pdf": This file is the coefficient plot in either PDF or PNG format.

-p PERCENT, --percent PERCENT
The maximum allowable percentage of missing CpGs. Set to 0.2 (20%) by default, which means that if more than 20% of the clock CpGs are missing, the estimation of DNAm age cannot be performed.

-d DELIMITER, --delimiter DELIMITER
Separator (usually TAB or comma) used in the input file. If the separator is not provided, the program will automatically detect the separator.

-f {pdf,png}, --format {pdf,png}
Figure format of the output coef plot. It must be "pdf" or "png". The default is "pdf".

-m meta_file, --metadata meta_file
This file contains the meta information for each sample. This file must have a header row, which contains the names or labels for variables (such as "Sex", "Age"). If the header row includes an "Age" column, a scatter plot will be generated to display the correlation between chronological age and predicted DNAm age. The first column of this file should contain sample IDs. The default value is None, indicating that no meta information file is provided.

-l log_file, --log log_file
This file is used to save the log information. By default, if no file is specified (None), the log information will be printed to the screen.

--impute {-1,0,1,2,3,4,5,6,7,8,9,10}
The imputation method code must be one of the 12 digits including (-1, 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10). The interpretations are: -1: Remove CpGs with any missing values. 0: Fill all missing values with '0.0'. 1: Fill all missing values with '1.0'. 2: Fill the missing values with ****column mean**** 3: Fill the missing values with ****column median**** 4: Fill the missing values with ****column min**** 5: Fill the missing values with ****column max**** 6: Fill the missing values with ****row mean**** 7: Fill the missing values with ****row median**** 8: Fill the missing values with ****row min**** 9: Fill the missing values with ****row max**** 10: Fill the missing values with ****external reference**** If 10 is specified, an external reference file must be

(continues on next page)

(continued from previous page)

```

        provided.
-r ref_file, --ref ref_file
    The external reference file contains two columns,
    separated by either tabs or commas. The first column
    represents the probe ID, while the second column
    contains the corresponding beta values.
--overwrite
    If set, over-write existing output files.
--debug
    If set, print detailed information for debugging.

```

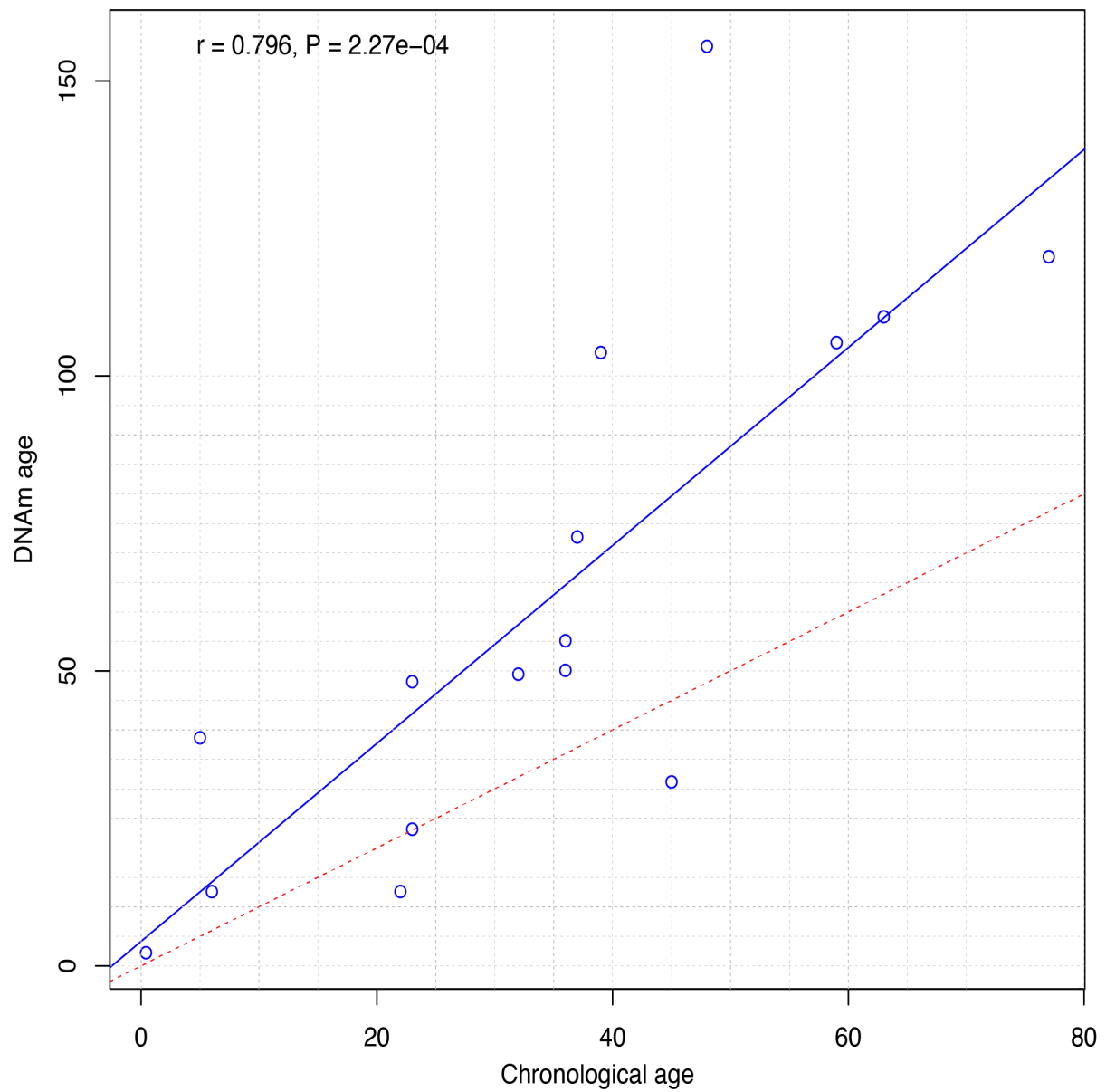
14.2 Example

```
$ epical Cortical Test4_brain_N16_EPICv2_beta.tsv.gz -m Test4_brain_N16_EPICv2_info.tsv
-o Cortical
```

```

2024-01-23 03:18:09 [INFO] The prefix of output files is set to "Cortical".
2024-01-23 03:18:09 [INFO] Loading Cortical clock data ...
2024-01-23 03:18:09 [INFO] Clock's name: "CorticalClock"
2024-01-23 03:18:09 [INFO] Clock was trained from: "brain cortex"
2024-01-23 03:18:09 [INFO] Clock's unit: "years"
2024-01-23 03:18:09 [INFO] Number of CpGs used: 347
...

```



Predictor CpGs	200
Unit	Year
Target Tissue	skeletal muscle
Target Population	Adult
Method	Elastic Net regression
Reference	Voisin S, et al, J Cachexia Sarcopenia Muscle. (2020).
Alias	

15.1 Usage

```
usage: epical MEAT [-h] [-o out_prefix] [-p PERCENT] [-d DELIMITER]
                  [-f {pdf,png}] [-m meta_file] [-l log_file]
                  [--impute {-1,0,1,2,3,4,5,6,7,8,9,10}] [-r ref_file]
                  [--overwrite] [--debug]
                  Input_file

positional arguments:
  Input_file            The input tabular structure file containing DNA
                        methylation data. This file must have a header row,
                        which contains the names or labels for samples. The
                        first column of this file should contain CpG IDs. The
                        remaining cells in the file should contain DNA
                        methylation beta values, represented as floating-
                        point numbers between 0 and 1. Use a TAB, comma, or any
                        other delimiter to separate the columns. Use 'NaN' or
                        'NA' to represent missing values. This file can be a
                        regular text file or compressed file (".gz", ".Z",
                        ".z", ".bz", ".bz2", ".bzip2").

options:
  -h, --help            show this help message and exit
  -o out_prefix, --output out_prefix
                        The PREFIX of output files. If no PREFIX is provided,
                        the default prefix "clock_name_out" is used. The
                        generated output files include:
                        "<PREFIX>.DNAm_age.tsv": This file contains the
                        predicted DNAm age. "<PREFIX>.used_CpGs.tsv": This
```

(continues on next page)

(continued from previous page)

file lists the CpGs that were used to calculate the DNAm age. "<PREFIX>.missed_CpGs.txt": This file provides a list of clock CpGs that were missed or excluded from the input file. "<PREFIX>.coef.tsv": This file contains a list of clock CpGs along with their coefficients. The last column indicates whether the CpG is included in the calculation.

"<PREFIX>.plots.R": This file is an R script used to generate visualization plots.

"<PREFIX>.coef_plot.pdf": This file is the coefficient plot in either PDF or PNG format.

-p PERCENT, --percent PERCENT
The maximum allowable percentage of missing CpGs. Set to 0.2 (20%) by default, which means that if more than 20% of the clock CpGs are missing, the estimation of DNAm age cannot be performed.

-d DELIMITER, --delimiter DELIMITER
Separator (usually TAB or comma) used in the input file. If the separator is not provided, the program will automatically detect the separator.

-f {pdf,png}, --format {pdf,png}
Figure format of the output coef plot. It must be "pdf" or "png". The default is "pdf".

-m meta_file, --metadata meta_file
This file contains the meta information for each sample. This file must have a header row, which contains the names or labels for variables (such as "Sex", "Age"). If the header row includes an "Age" column, a scatter plot will be generated to display the correlation between chronological age and predicted DNAm age. The first column of this file should contain sample IDs. The default value is None, indicating that no meta information file is provided.

-l log_file, --log log_file
This file is used to save the log information. By default, if no file is specified (None), the log information will be printed to the screen.

--impute {-1,0,1,2,3,4,5,6,7,8,9,10}
The imputation method code must be one of the 12 digits including (-1, 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10). The interpretations are: -1: Remove CpGs with any missing values. 0: Fill all missing values with '0.0'. 1: Fill all missing values with '1.0'. 2: Fill the missing values with ****column mean**** 3: Fill the missing values with ****column median**** 4: Fill the missing values with ****column min**** 5: Fill the missing values with ****column max**** 6: Fill the missing values with ****row mean**** 7: Fill the missing values with ****row median**** 8: Fill the missing values with ****row min**** 9: Fill the missing values with ****row max**** 10: Fill the missing values with ****external reference**** If 10 is specified, an external reference file must be

(continues on next page)

(continued from previous page)

```

        provided.
-r ref_file, --ref ref_file
    The external reference file contains two columns,
    separated by either tabs or commas. The first column
    represents the probe ID, while the second column
    contains the corresponding beta values.
--overwrite
    If set, over-write existing output files.
--debug
    If set, print detailed information for debugging.

```

15.2 Example

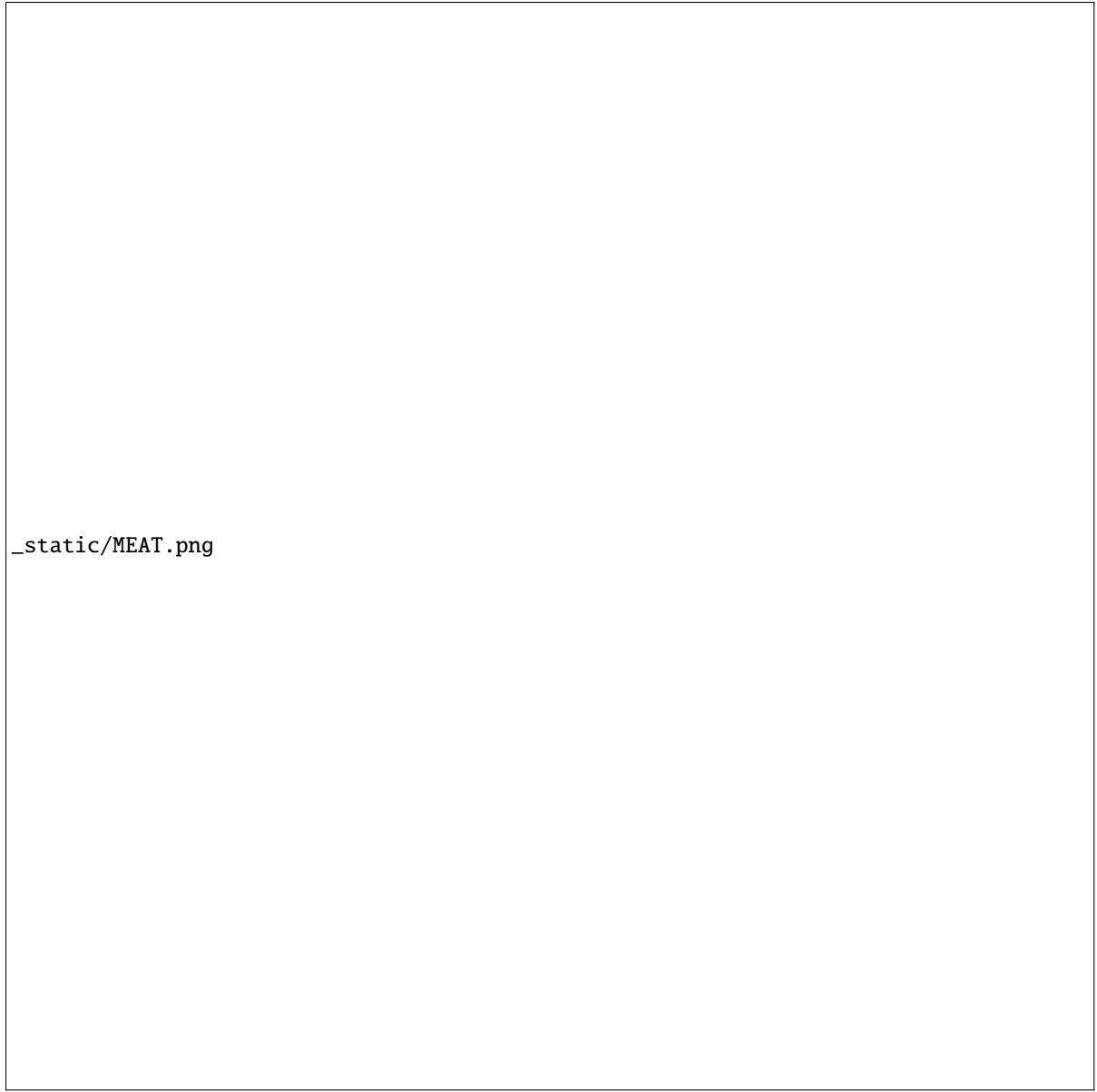
run muscle clock on brain samples...

```
$ epical MEAT Test4_brain_N16_EPICv2_beta.tsv.gz -m Test4_brain_N16_EPICv2_info.tsv -o MEAT
```

```

2024-04-25 11:40:25 [INFO] The prefix of output files is set to "MEAT".
2024-04-25 11:40:25 [INFO] Loading MEAT clock data ...
2024-04-25 11:40:25 [INFO] Clock's name: "MuscleClock"
2024-04-25 11:40:25 [INFO] Clock was trained from: "muscle"
2024-04-25 11:40:25 [INFO] Clock's unit: "years"
2024-04-25 11:40:25 [INFO] Number of CpGs used: 200
...

```



`_static/MEAT.png`

16.1 Description

1. EPM is a modeling tool, not a epigenetic clock. It can be used to model any phenotypes including aging.
2. In contrast to regression bases methods, the EPM does not assume a linear relationship between the epigenetic state and a trait of interest.
3. In contrast to other clocks within this package, the EPM algorithm does NOT come with pre-selected CpGs and pre-calculated coefficients. Instead, users are required to provide training data to build the model. In essence, EPM performs CpG selection, model fitting, and prediction simultaneously. For more technical details, refer to the [EPM tutorial](#).
4. The `epical EPM` command need two input files: the first being the beta value spreadsheet, and the second being the meta-information file that specifies the training and test samples (see the example below).

Predictor CpGs	Subject to training data
Unit	Subject to training data
Target Tissue	Subject to training data
Target Population	Subject to training data
Method	Fast conditional expectation maximization
Reference	Farrell C, et al, Bioinformatics. (2020).
Alias	EpigeneticPacemaker (EPM)

16.2 Usage

```
usage: epical EPM [-h] [-o out_prefix] [-p PCC] [-n NITER] [-k KFOLD]
                  [-e ETOL] [-d DELIMITER] [-f {pdf,png}] [-l log_file]
                  [-i {-1,0,1,2,3,4,5,6,7,8,9,10}] [-r ref_file] [--debug]
                  Input_file meta_file
```

positional arguments:

Input_file The input tabular structure file containing DNA methylation data. This file must have a header row, which contains the names or labels for samples. The first column of this file should contain CpG IDs. The remaining cells in the file should contain DNA methylation beta values, represented as floating-point numbers between 0 and 1. Use a TAB, comma, or any

(continues on next page)

(continued from previous page)

```

other delimiter to separatethe columns. Use 'NaN' or
'NA' to represent missing values. This file can be a
regular text file or compressed file (".gz", ".Z",
".z", ".bz", ".bz2", ".bzip2").
meta_file      This file contains the meta information for each
                sample. This file must have a header row, which
                contains the names or labels for variables. The 'Age'
                variable must exist. The 'Designation' variable is
                used to designate training and testing samples.

options:
-h, --help      show this help message and exit
-o out_prefix, --output out_prefix
                The PREFIX of output files. If no PREFIX is provided,
                the default prefix "EPM_out" is used. The generated
                output files include: "<PREFIX>.test_EPM_age.tsv": The
                predicted EPM age for testing samples.
                "<PREFIX>.train_EPM_age.tsv": The predicted EPM age
                for training samples. "<PREFIX>.test_EPM_age.pdf" or
                "<PREFIX>.test_EPM_age.png": Scatter plot showing the
                trend between the predicted EPM ages and chronological
                ages for testing samples. "<PREFIX>.train_EPM_age.pdf"
                or "<PREFIX>.train_EPM_age.png": Scatter plot showing
                the trend between the predicted EPM ages and
                chronological ages for training samples.
                "<PREFIX>.test_selected_CpGs.tsv": Selected feature
                CpGs and their beta values for testing samples.
                "<PREFIX>.train_selected_CpGs.tsv": Selected feature
                CpGs and their beta values for training samples.
-p PCC, --pcc PCC  Threshold of absolute Pearson correlation coefficient
                between chronological age and beta values. This cutoff
                is used to select age-associated CpG sites.
-n NITER, --niter NITER
                Iteration times of expectation-maximization.
-k KFOLD, --kfold KFOLD
                Folds for cross-valuation.
-e ETOL, --etol ETOL  Error tolerance during model fitting. The acceptable
                level of deviation between the EPM predicted age and
                the chronological age.
-d DELIMITER, --delimiter DELIMITER
                Separator (usually TAB or comma) used in the input
                file. If the separator is not provided, the program
                will automatically detect the separator.
-f {pdf,png}, --format {pdf,png}
                Figure format of the output coef plot. It must be
                "pdf" or "png". The default is "pdf".
-l log_file, --log log_file
                This file is used to save the log information. By
                default, if no file is specified (None), the log
                information will be printed to the screen.
-i {-1,0,1,2,3,4,5,6,7,8,9,10}, --impute {-1,0,1,2,3,4,5,6,7,8,9,10}
                The imputation method code must be one of the 12

```

(continues on next page)

(continued from previous page)

```

digits including (-1, 0, 1, 2, 3, 4, 5, 6, 7, 8, 9,
10). The interpretations are: -1: Remove CpGs with any
missing values. 0: Fill all missing values with '0.0'.
1: Fill all missing values with '1.0'. 2: Fill the
missing values with **column mean** 3: Fill the
missing values with **column median** 4: Fill the
missing values with **column min** 5: Fill the missing
values with **column max** 6: Fill the missing values
with **row mean** 7: Fill the missing values with
**row median** 8: Fill the missing values with **row
min** 9: Fill the missing values with **row max** 10:
Fill the missing values with **external reference** If
10 is specified, an external reference file must be
provided.
-r ref_file, --ref ref_file
    The external reference file contains two columns,
    separated by either tabs or commas. The first column
    represents the probe ID, while the second column
    contains the corresponding beta values.
--debug
    If set, print detailed information for debugging.

```

16.3 Example

Prepare the meta-information file. In this instance, the 80 samples were shuffled, and subsequently, 60 (75%) samples were chosen as “training” samples, while the remaining 20 (25%) samples were designated as “testing” samples.

Note: The meta-information file must have two columns named as “Age” and “Designation”.

```
$ cat Test2_blood_N80_450K_info2.tsv
```

```

ID  Sex Age Designation
I304 F 37  train
I080 M 12  train
I513 M 41  train
I118 M 10  train
I063 F 13  train
I550 M 19  train
I097 M 44  train
I282 M 50  train
I257 M 48  train
I015 M 24  train
I302 F 43  train
I290 M 14  train
I224 F 37  train
I057 F 44  train
I255 F 17  train
I572 F 41  train
I028 F 13  train
I215 F 16  train

```

(continues on next page)

(continued from previous page)

I007	M	20	train
I537	F	21	train
I060	F	20	train
I481	F	56	train
I037	M	20	train
I251	F	48	train
I163	F	11	train
I589	M	59	train
I109	M	49	train
I006	M	53	train
I167	M	53	train
I004	F	11	train
I148	M	52	train
I482	F	24	train
I270	F	17	train
I074	F	12	train
I159	M	40	train
I478	M	39	train
I286	F	17	train
I127	F	15	train
I072	M	12	train
I035	F	19	train
I228	F	55	train
I020	M	11	train
I292	F	14	train
I114	M	47	train
I229	M	21	train
I349	M	43	train
I173	F	47	train
I032	M	13	train
I013	M	10	train
I495	M	46	train
I107	M	18	train
I213	F	16	train
I190	F	45	train
I058	M	19	train
I104	F	10	train
I211	M	16	train
I135	M	39	train
I091	F	40	train
I311	F	58	train
I152	F	42	train
I280	F	14	test
I182	M	51	test
I092	M	15	test
I391	F	52	test
I458	M	54	test
I279	M	55	test
I066	M	15	test
I077	F	42	test
I234	F	51	test
I024	F	49	test

(continues on next page)

(continued from previous page)

```

I603 F 18 test
I379 M 57 test
I418 F 46 test
I268 M 38 test
I338 M 18 test
I034 F 38 test
I230 F 50 test
I412 M 60 test
I188 M 45 test
I274 F 54 test

```

```
$ epical EPM Test2_blood_N80_450K_beta.tsv.gz Test2_blood_N80_450K_info.tsv -o EPM
```

```

2024-01-23 07:26:09 [INFO] The prefix of output files is set to "EPM".
2024-01-23 07:26:09 [INFO] Read input beta file: "Test2_blood_N80_450K_beta.tsv.gz"
2024-01-23 07:26:39 [INFO] Fill missing values with ZEROs ...
2024-01-23 07:26:39 [INFO] Input file: "Test2_blood_N80_450K_beta.tsv.gz", Number of
↳CpGs: 478313, Number of samples: 80
2024-01-23 07:26:39 [INFO] Read meta information file: "Test2_blood_N80_450K_info2.tsv"
2024-01-23 07:26:39 [INFO] Split samples into training and testing sets ...
2024-01-23 07:26:39 [INFO] 60 samples are included in training set: I304, I080, I513,
↳I118, I063 ...
2024-01-23 07:26:39 [INFO] 20 samples are included in testing set: I280, I182, I092,
↳I391, I458 ...
2024-01-23 07:26:39 [INFO] Calculate pearson correlation coefficients ...
2024-01-23 07:26:40 [INFO] 100 CpG sites are selected: cg00103778, cg00303541,
↳cg00329615, cg00439658, cg00462994 ...
2024-01-23 07:26:40 [INFO] Save beta values of selected CpGs to "EPM.train_selected_
↳CpGs.tsv"
2024-01-23 07:26:40 [INFO] Save beta values of selected CpGs to "EPM.test_selected_CpGs.
↳tsv"
2024-01-23 07:26:40 [INFO] Initialize the EPM model ...
2024-01-23 07:26:40 [INFO] Fit the EPM model using training data ...
2024-01-23 07:26:40 [INFO] Get training sample EPM predictions (when left out) ...
2024-01-23 07:26:40 [INFO] Save predicted EPM ages of training samples to "EPM.train_EPM_
↳age.tsv"
2024-01-23 07:26:40 [INFO] Predict testing samples ...
2024-01-23 07:26:40 [INFO] Save predicted EPM age of testing samples to "EPM.test_EPM_
↳age.tsv"
2024-01-23 07:26:40 [INFO] Generate scatter plot of test samples and save to "EPM.test_
↳EPM_age.pdf"

```

16.4 Ouput

A total of six files were generated.

1. EPM.train_selected_CpGs.tsv

Selected predictor CpGs and their beta values from training samples.

2. EPM.train_EPM_age.tsv

Predicted age for training samples. Please note, the “cross validated version of the EPM” is implemented here, so **prediction is made for each sample when the sample is left out of model fitting**.

3. EPM.train_EPM_age.pdf

Scatter plot showing the concordance between “Chronological age” and “predicted age” (for training samples).

4. EPM.test_selected_CpGs.tsv

Selected predictor CpGs and their beta values from testing samples.

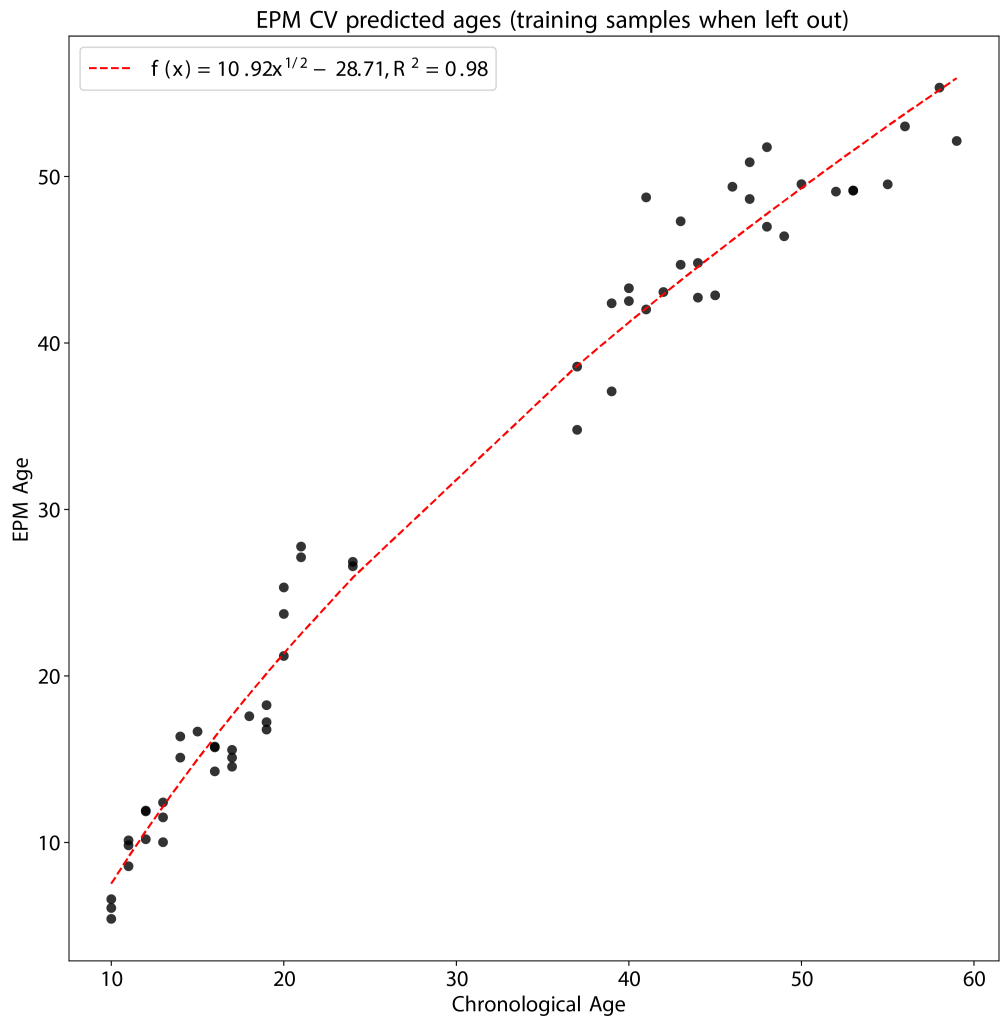
5. EPM.test_EPM_age.tsv

Predicted age for testing samples.

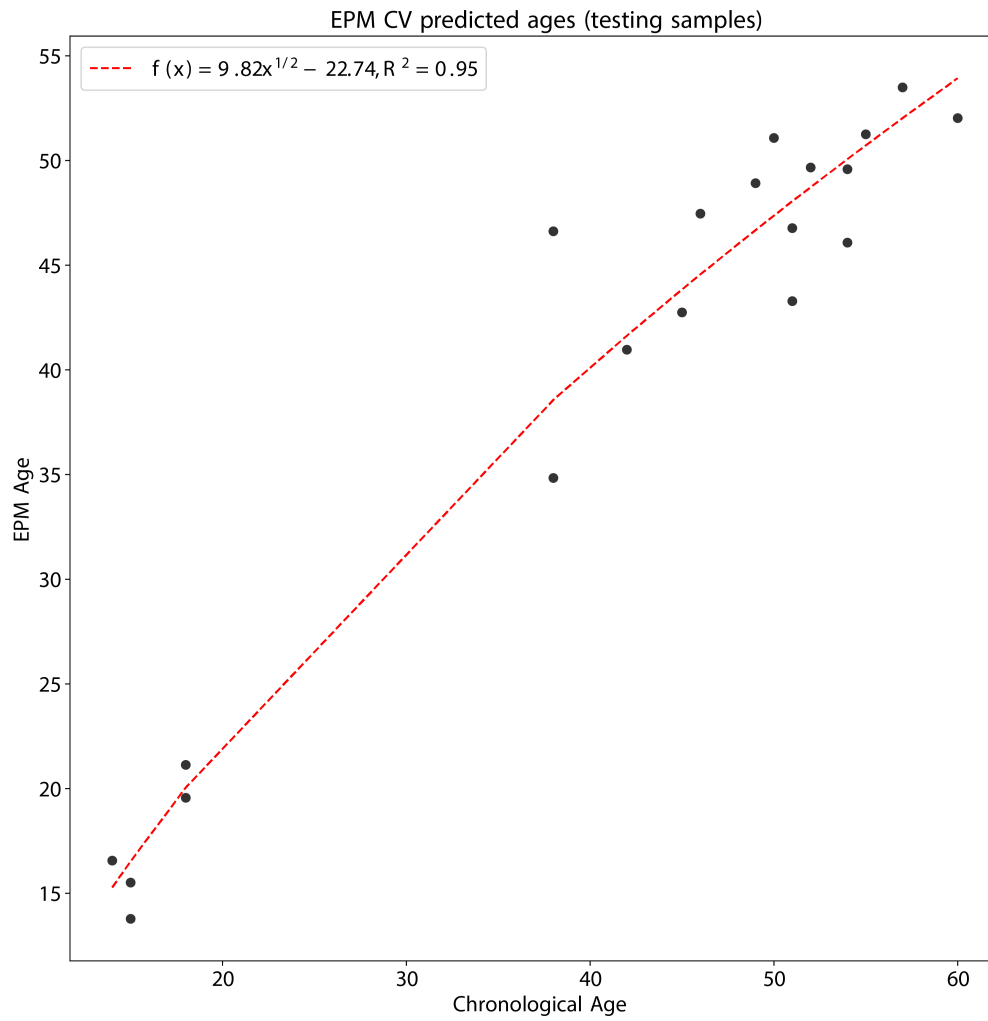
6. EPM.test_EPM_age.pdf

Scatter plot showing the concordance between “Chronological age” and “predicted age” (for testing samples).

EPM_train.png



EPM_test.png



LU_DNAMTL

Predictor CpGs	140
Unit	Kb
Target Tissue	Blood
Target Population	Adult
Method	Elastic Net regression
Reference	Lu, Ake, et al., Aging. (2019).
Alias	

17.1 Usage

```
usage: epical Lu_DNAMTL [-h] [-o out_prefix] [-p PERCENT] [-d DELIMITER]
                        [-f {pdf,png}] [-m meta_file] [-l log_file]
                        [--impute {-1,0,1,2,3,4,5,6,7,8,9,10}] [-r ref_file]
                        [--debug] [--overwrite]
                        Input_file

positional arguments:
  Input_file            The input tabular structure file containing DNA
                        methylation data. This file must have a header row,
                        which contains the names or labels for samples. The
                        first column of this file should contain CpG IDs. The
                        remaining cells in the file should contain DNA
                        methylation beta values, represented as floating-
                        point numbers between 0 and 1. Use a TAB, comma, or any
                        other delimiter to separate the columns. Use 'NaN' or
                        'NA' to represent missing values. This file can be a
                        regular text file or compressed file (".gz", ".Z",
                        ".z", ".bz", ".bz2", ".bzip2").

options:
  -h, --help            show this help message and exit
  -o out_prefix, --output out_prefix
                        The PREFIX of output files. If no PREFIX is provided,
                        the default prefix "clock_name_out" is used. The
                        generated output files include:
                        "<PREFIX>.DNAm_age.tsv": This file contains the
                        predicted DNAm age. "<PREFIX>.used_CpGs.tsv": This
```

(continues on next page)

(continued from previous page)

file lists the CpGs that were used to calculate the DNAm age. "<PREFIX>.missed_CpGs.txt": This file provides a list of clock CpGs that were missed or excluded from the input file. "<PREFIX>.coef.tsv": This file contains a list of clock CpGs along with their coefficients. The last column indicates whether the CpG is included in the calculation.

"<PREFIX>.plots.R": This file is an R script used to generate visualization plots.

"<PREFIX>.coef_plot.pdf": This file is the coefficient plot in either PDF or PNG format.

-p PERCENT, --percent PERCENT
The maximum allowable percentage of missing CpGs. Set to 0.2 (20%) by default, which means that if more than 20% of the clock CpGs are missing, the estimation of DNAm age cannot be performed.

-d DELIMITER, --delimiter DELIMITER
Separator (usually TAB or comma) used in the input file. If the separator is not provided, the program will automatically detect the separator.

-f {pdf,png}, --format {pdf,png}
Figure format of the output coef plot. It must be "pdf" or "png". The default is "pdf".

-m meta_file, --metadata meta_file
This file contains the meta information for each sample. This file must have a header row, which contains the names or labels for variables (such as "Sex", "Age"). If the header row includes an "Age" column, a scatter plot will be generated to display the correlation between chronological age and predicted DNAm age. The first column of this file should contain sample IDs. The default value is None, indicating that no meta information file is provided.

-l log_file, --log log_file
This file is used to save the log information. By default, if no file is specified (None), the log information will be printed to the screen.

--impute {-1,0,1,2,3,4,5,6,7,8,9,10}
The imputation method code must be one of the 12 digits including (-1, 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10). The interpretations are: -1: Remove CpGs with any missing values. 0: Fill all missing values with '0.0'. 1: Fill all missing values with '1.0'. 2: Fill the missing values with **column mean** 3: Fill the missing values with **column median** 4: Fill the missing values with **column min** 5: Fill the missing values with **column max** 6: Fill the missing values with **row mean** 7: Fill the missing values with **row median** 8: Fill the missing values with **row min** 9: Fill the missing values with **row max** 10: Fill the missing values with **external reference** If 10 is specified, an external reference file must be

(continues on next page)

(continued from previous page)

```

        provided.
-r ref_file, --ref ref_file
    The external reference file contains two columns,
    separated by either tabs or commas. The first column
    represents the probe ID, while the second column
    contains the corresponding beta values.
--debug
    Print detailed information for debugging.
--overwrite
    If set, print detailed information for debugging.

```

17.2 Example

```

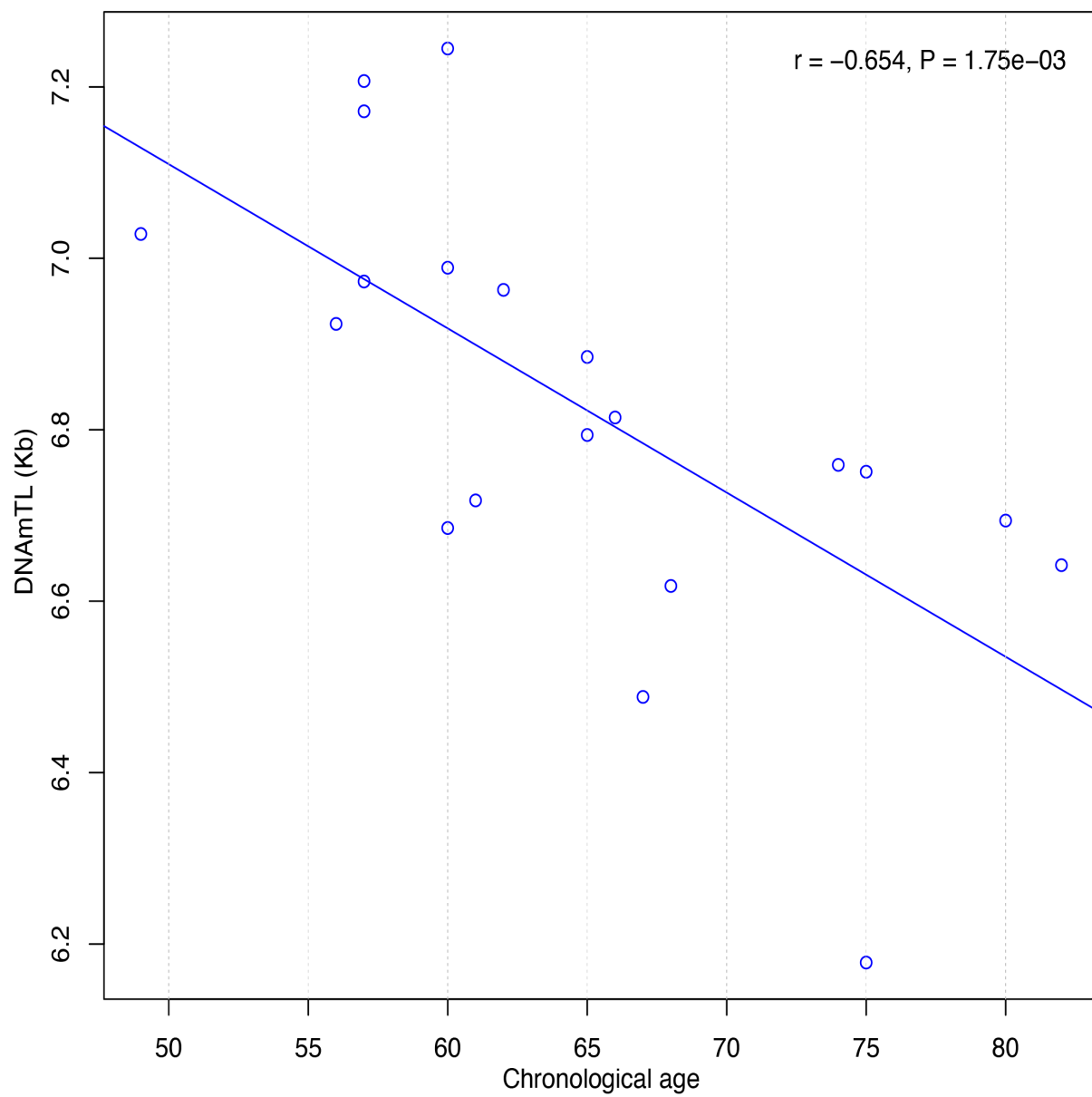
$ epical Lu_DNAmtL Test1_blood_N20_EPICv1_beta.tsv.gz -m Test1_blood_N20_EPICv1_info.tsv
-o Lu_DNAmtL

```

```

2024-01-17 10:01:55 [INFO] The prefix of output files is set to "Lu_DNAmtL".
2024-01-17 10:01:55 [INFO] Loading Lu_DNAmtL clock data ...
2024-01-17 10:01:55 [INFO] Clock's name: "Lu_DNAmtL"
2024-01-17 10:01:55 [INFO] Clock was trained from: "blood"
2024-01-17 10:01:55 [INFO] Clock's unit: "Kilobase"
2024-01-17 10:01:55 [INFO] Number of CpGs used: 140
...

```



Predictor CpGs	111
Unit	Year
Target Tissue	Blood, Saliva
Target Population	Children (9 to 17)
Method	Elastic Net regression
Reference	Wu X, et al., Aging. (2019).
Alias	

18.1 Usage

```
usage: epical Ped_Wu [-h] [-o out_prefix] [-p PERCENT] [-d DELIMITER]
                    [-f {pdf,png}] [-m meta_file] [-l log_file]
                    [--impute {-1,0,1,2,3,4,5,6,7,8,9,10}] [-r ref_file]
                    [--debug] [--overwrite]
                    Input_file

positional arguments:
  Input_file            The input tabular structure file containing DNA
                        methylation data. This file must have a header row,
                        which contains the names or labels for samples. The
                        first column of this file should contain CpG IDs. The
                        remaining cells in the file should contain DNA
                        methylation beta values, represented as floating-
                        point numbers between 0 and 1. Use a TAB, comma, or any
                        other delimiter to separate the columns. Use 'NaN' or
                        'NA' to represent missing values. This file can be a
                        regular text file or compressed file (".gz", ".Z",
                        ".z", ".bz", ".bz2", ".bzip2").

options:
  -h, --help            show this help message and exit
  -o out_prefix, --output out_prefix
                        The PREFIX of output files. If no PREFIX is provided,
                        the default prefix "clock_name_out" is used. The
                        generated output files include:
                        "<PREFIX>.DNAm_age.tsv": This file contains the
                        predicted DNAm age. "<PREFIX>.used_CpGs.tsv": This
```

(continues on next page)

(continued from previous page)

file lists the CpGs that were used to calculate the DNAm age. "<PREFIX>.missed_CpGs.txt": This file provides a list of clock CpGs that were missed or excluded from the input file. "<PREFIX>.coef.tsv": This file contains a list of clock CpGs along with their coefficients. The last column indicates whether the CpG is included in the calculation.

"<PREFIX>.plots.R": This file is an R script used to generate visualization plots.

"<PREFIX>.coef_plot.pdf": This file is the coefficient plot in either PDF or PNG format.

-p PERCENT, --percent PERCENT
The maximum allowable percentage of missing CpGs. Set to 0.2 (20%) by default, which means that if more than 20% of the clock CpGs are missing, the estimation of DNAm age cannot be performed.

-d DELIMITER, --delimiter DELIMITER
Separator (usually TAB or comma) used in the input file. If the separator is not provided, the program will automatically detect the separator.

-f {pdf,png}, --format {pdf,png}
Figure format of the output coef plot. It must be "pdf" or "png". The default is "pdf".

-m meta_file, --metadata meta_file
This file contains the meta information for each sample. This file must have a header row, which contains the names or labels for variables (such as "Sex", "Age"). If the header row includes an "Age" column, a scatter plot will be generated to display the correlation between chronological age and predicted DNAm age. The first column of this file should contain sample IDs. The default value is None, indicating that no meta information file is provided.

-l log_file, --log log_file
This file is used to save the log information. By default, if no file is specified (None), the log information will be printed to the screen.

--impute {-1,0,1,2,3,4,5,6,7,8,9,10}
The imputation method code must be one of the 12 digits including (-1, 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10). The interpretations are: -1: Remove CpGs with any missing values. 0: Fill all missing values with '0.0'. 1: Fill all missing values with '1.0'. 2: Fill the missing values with **column mean** 3: Fill the missing values with **column median** 4: Fill the missing values with **column min** 5: Fill the missing values with **column max** 6: Fill the missing values with **row mean** 7: Fill the missing values with **row median** 8: Fill the missing values with **row min** 9: Fill the missing values with **row max** 10: Fill the missing values with **external reference** If 10 is specified, an external reference file must be

(continues on next page)

(continued from previous page)

```

        provided.
-r ref_file, --ref ref_file
    The external reference file contains two columns,
    separated by either tabs or commas. The first column
    represents the probe ID, while the second column
    contains the corresponding beta values.
--debug
    If set, print detailed information for debugging.
--overwrite
    If set, over-write existing output files.

```

18.2 Example

```
$ epical Ped_Wu Test2_blood_N80_450K_beta.tsv.gz -m Test2_blood_N80_450K_info.tsv -o
PedWu_out
```

```

2024-01-22 10:08:01 [INFO] The prefix of output files is set to "PedWu_out".
2024-01-22 10:08:01 [INFO] Loading Ped_Wu clock data ...
2024-01-22 10:08:01 [INFO] Clock's name: "Wu_Children"
2024-01-22 10:08:01 [INFO] Clock was trained from: "blood,saliva"
2024-01-22 10:08:01 [INFO] Clock's unit: "years"
2024-01-22 10:08:01 [INFO] Number of CpGs used: 111
...

```


Predictor CpGs	94
Unit	Year
Target Tissue	Buccal cells
Target Population	Children (0 to 20)
Method	Elastic Net regression
Reference	McEwen LM, et al., Proc Natl Acad Sci U S A. (2020).
Alias	

19.1 Usage

```
usage: epical PedBE [-h] [-o out_prefix] [-p PERCENT] [-d DELIMITER]
                  [-f {pdf,png}] [-m meta_file] [-l log_file]
                  [--impute {-1,0,1,2,3,4,5,6,7,8,9,10}] [-r ref_file]
                  [--debug] [--overwrite]
                  Input_file

positional arguments:
  Input_file            The input tabular structure file containing DNA
                        methylation data. This file must have a header row,
                        which contains the names or labels for samples. The
                        first column of this file should contain CpG IDs. The
                        remaining cells in the file should contain DNA
                        methylation beta values, represented as floating-
                        point numbers between 0 and 1. Use a TAB, comma, or any
                        other delimiter to separate the columns. Use 'NaN' or
                        'NA' to represent missing values. This file can be a
                        regular text file or compressed file (".gz", ".Z",
                        ".z", ".bz", ".bz2", ".bzip2").

options:
  -h, --help            show this help message and exit
  -o out_prefix, --output out_prefix
                        The PREFIX of output files. If no PREFIX is provided,
                        the default prefix "clock_name_out" is used. The
                        generated output files include:
                        "<PREFIX>.DNAm_age.tsv": This file contains the
                        predicted DNAm age. "<PREFIX>.used_CpGs.tsv": This
```

(continues on next page)

(continued from previous page)

file lists the CpGs that were used to calculate the DNAm age. "<PREFIX>.missed_CpGs.txt": This file provides a list of clock CpGs that were missed or excluded from the input file. "<PREFIX>.coef.tsv": This file contains a list of clock CpGs along with their coefficients. The last column indicates whether the CpG is included in the calculation.

"<PREFIX>.plots.R": This file is an R script used to generate visualization plots.

"<PREFIX>.coef_plot.pdf": This file is the coefficient plot in either PDF or PNG format.

-p PERCENT, --percent PERCENT
The maximum allowable percentage of missing CpGs. Set to 0.2 (20%) by default, which means that if more than 20% of the clock CpGs are missing, the estimation of DNAm age cannot be performed.

-d DELIMITER, --delimiter DELIMITER
Separator (usually TAB or comma) used in the input file. If the separator is not provided, the program will automatically detect the separator.

-f {pdf,png}, --format {pdf,png}
Figure format of the output coef plot. It must be "pdf" or "png". The default is "pdf".

-m meta_file, --metadata meta_file
This file contains the meta information for each sample. This file must have a header row, which contains the names or labels for variables (such as "Sex", "Age"). If the header row includes an "Age" column, a scatter plot will be generated to display the correlation between chronological age and predicted DNAm age. The first column of this file should contain sample IDs. The default value is None, indicating that no meta information file is provided.

-l log_file, --log log_file
This file is used to save the log information. By default, if no file is specified (None), the log information will be printed to the screen.

--impute {-1,0,1,2,3,4,5,6,7,8,9,10}
The imputation method code must be one of the 12 digits including (-1, 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10). The interpretations are: -1: Remove CpGs with any missing values. 0: Fill all missing values with '0.0'. 1: Fill all missing values with '1.0'. 2: Fill the missing values with ****column mean**** 3: Fill the missing values with ****column median**** 4: Fill the missing values with ****column min**** 5: Fill the missing values with ****column max**** 6: Fill the missing values with ****row mean**** 7: Fill the missing values with ****row median**** 8: Fill the missing values with ****row min**** 9: Fill the missing values with ****row max**** 10: Fill the missing values with ****external reference**** If 10 is specified, an external reference file must be

(continues on next page)

(continued from previous page)

```

        provided.
-r ref_file, --ref ref_file
    The external reference file contains two columns,
    separated by either tabs or commas. The first column
    represents the probe ID, while the second column
    contains the corresponding beta values.
--debug
    If set, print detailed information for debugging.
--overwrite
    If set, over-write existing output files.

```

19.2 Example

```
$ epical PedBE Test2_blood_N80_450K_beta.tsv.gz -m Test2_blood_N80_450K_info.tsv -o
PedBE_out
```

```

2024-01-22 11:33:09 [INFO] The prefix of output files is set to "PedBE_out".
2024-01-22 11:33:09 [INFO] Loading PedBE clock data ...
2024-01-22 11:33:09 [INFO] Clock's name: "McEwen_PedBE"
2024-01-22 11:33:09 [INFO] Clock was trained from: "buccal cells"
2024-01-22 11:33:09 [INFO] Clock's unit: "years"
2024-01-22 11:33:09 [INFO] Number of CpGs used: 94
...

```


Predictor CpGs	96
Unit	Day
Target Tissue	Cord blood
Target Population	New born (Gestational age)
Method	Lasso regression
Reference	Bohlin J, et al., Genome Biol. (2016).
Alias	

20.1 Usage

```
usage: epical GA_Bohlin [-h] [-o out_prefix] [-p PERCENT] [-d DELIMITER]
                        [-f {pdf,png}] [-m meta_file] [-l mog_file]
                        [--impute {-1,0,1,2,3,4,5,6,7,8,9,10}] [-r ref_file]
                        [--debug] [--overwrite]
                        Input_file

positional arguments:
  Input_file            The input tabular structure file containing DNA
                        methylation data. This file must have a header row,
                        which contains the names or labels for samples. The
                        first column of this file should contain CpG IDs. The
                        remaining cells in the file should contain DNA
                        methylation beta values, represented as floating-
                        point numbers between 0 and 1. Use a TAB, comma, or any
                        other delimiter to separate the columns. Use 'NaN' or
                        'NA' to represent missing values. This file can be a
                        regular text file or compressed file (".gz", ".Z",
                        ".z", ".bz", ".bz2", ".bzip2").

options:
  -h, --help            show this help message and exit
  -o out_prefix, --output out_prefix
                        The PREFIX of output files. If no PREFIX is provided,
                        the default prefix "clock_name_out" is used. The
                        generated output files include:
                        "<PREFIX>.DNAm_age.tsv": This file contains the
                        predicted DNAm age. "<PREFIX>.used_CpGs.tsv": This
```

(continues on next page)

(continued from previous page)

file lists the CpGs that were used to calculate the DNAm age. "<PREFIX>.missed_CpGs.txt": This file provides a list of clock CpGs that were missed or excluded from the input file. "<PREFIX>.coef.tsv": This file contains a list of clock CpGs along with their coefficients. The last column indicates whether the CpG is included in the calculation. "<PREFIX>.plots.R": This file is an R script used to generate visualization plots. "<PREFIX>.coef_plot.pdf": This file is the coefficient plot in either PDF or PNG format.

-p PERCENT, --percent PERCENT
The maximum allowable percentage of missing CpGs. Set to 0.2 (20%) by default, which means that if more than 20% of the clock CpGs are missing, the estimation of DNAm age cannot be performed.

-d DELIMITER, --delimiter DELIMITER
Separator (usually TAB or comma) used in the input file. If the separator is not provided, the program will automatically detect the separator.

-f {pdf,png}, --format {pdf,png}
Figure format of the output coef plot. It must be "pdf" or "png". The default is "pdf".

-m meta_file, --metadata meta_file
This file contains the meta information for each sample. This file must have a header row, which contains the names or labels for variables (such as "Sex", "Age"). If the header row includes an "Age" column, a scatter plot will be generated to display the correlation between chronological age and predicted DNAm age. The first column of this file should contain sample IDs. The default value is None, indicating that no meta information file is provided.

-l mog_file, --log mog_file
This file is used to save the log information. By default, if no file is specified (None), the log information will be printed to the screen.

--impute {-1,0,1,2,3,4,5,6,7,8,9,10}
The imputation method code must be one of the 12 digits including (-1, 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10). The interpretations are: -1: Remove CpGs with any missing values. 0: Fill all missing values with '0.0'. 1: Fill all missing values with '1.0'. 2: Fill the missing values with ****column mean**** 3: Fill the missing values with ****column median**** 4: Fill the missing values with ****column min**** 5: Fill the missing values with ****column max**** 6: Fill the missing values with ****row mean**** 7: Fill the missing values with ****row median**** 8: Fill the missing values with ****row min**** 9: Fill the missing values with ****row max**** 10: Fill the missing values with ****external reference**** If 10 is specified, an external reference file must be

(continues on next page)

(continued from previous page)

```

        provided.
-r ref_file, --ref ref_file
    The external reference file contains two columns,
    separated by either tabs or commas. The first column
    represents the probe ID, while the second column
    contains the corresponding beta values.
--debug
    If set, print detailed information for debugging.
--overwrite
    If set, over-write existing output files.

```

20.2 Example

```

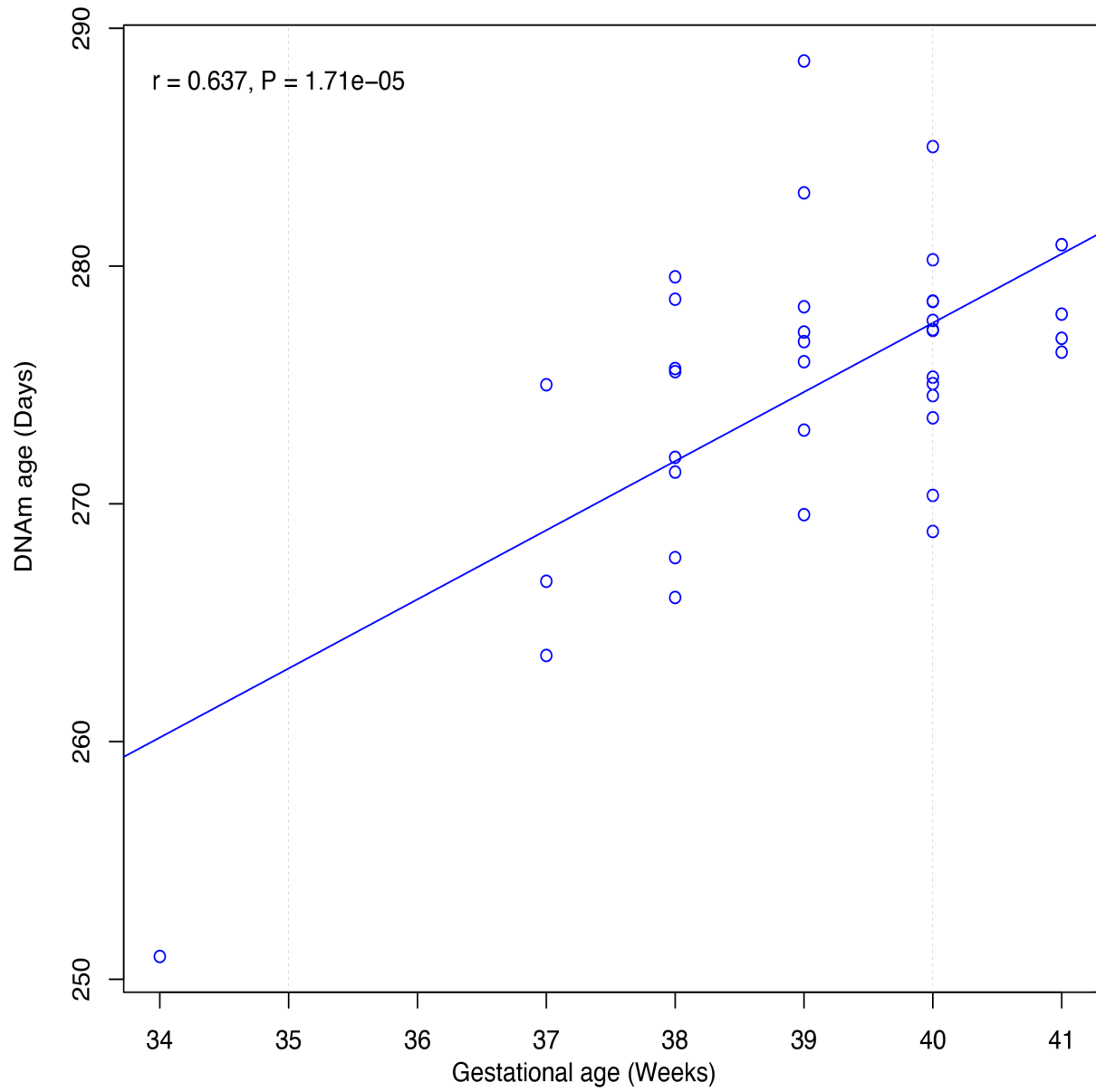
$ epical GA_Bohlin Test5_CordBlood_N38_450K_beta.tsv.gz -m Test5_CordBlood_N38_450K_info.
tsv -o GA_Bohlin --overwrite

```

```

2024-01-22 09:49:43 [INFO] The prefix of output files is set to "GA_Bohlin".
2024-01-22 09:49:43 [WARNING] Over write existing files with prefix: GA_Bohlin
2024-01-22 09:49:43 [INFO] Loading GA_Bohlin clock data ...
2024-01-22 09:49:43 [INFO] Clock's name: "Bohlin_gestational"
2024-01-22 09:49:43 [INFO] Clock was trained from: "cord blood"
2024-01-22 09:49:43 [INFO] Clock's unit: "days"
2024-01-22 09:49:43 [INFO] Number of CpGs used: 96
...

```



GA_HAFTORN

Predictor CpGs	176
Unit	Day
Target Tissue	Cord blood
Target Population	New born (Gestational age)
Method	Lasso regression
Reference	Haftorn KL., et al., Clinical epigenetics (2021) .
Alias	

21.1 Usage

```
usage: epical GA_Haftorn [-h] [-o out_prefix] [-p PERCENT] [-d DELIMITER]
                        [-f {pdf,png}] [-m meta_file] [-l mog_file]
                        [--impute {-1,0,1,2,3,4,5,6,7,8,9,10}] [-r ref_file]
                        [--debug] [--overwrite]
                        Input_file

positional arguments:
  Input_file            The input tabular structure file containing DNA
                        methylation data. This file must have a header row,
                        which contains the names or labels for samples. The
                        first column of this file should contain CpG IDs. The
                        remaining cells in the file should contain DNA
                        methylation beta values, represented as floating-
                        point numbers between 0 and 1. Use a TAB, comma, or any
                        other delimiter to separate the columns. Use 'NaN' or
                        'NA' to represent missing values. This file can be a
                        regular text file or compressed file (".gz", ".Z",
                        ".z", ".bz", ".bz2", ".bzip2").

options:
  -h, --help            show this help message and exit
  -o out_prefix, --output out_prefix
                        The PREFIX of output files. If no PREFIX is provided,
                        the default prefix "clock_name_out" is used. The
                        generated output files include:
                        "<PREFIX>.DNAm_age.tsv": This file contains the
                        predicted DNAm age. "<PREFIX>.used_CpGs.tsv": This
```

(continues on next page)

(continued from previous page)

file lists the CpGs that were used to calculate the DNAm age. "<PREFIX>.missed_CpGs.txt": This file provides a list of clock CpGs that were missed or excluded from the input file. "<PREFIX>.coef.tsv": This file contains a list of clock CpGs along with their coefficients. The last column indicates whether the CpG is included in the calculation.

"<PREFIX>.plots.R": This file is an R script used to generate visualization plots.

"<PREFIX>.coef_plot.pdf": This file is the coefficient plot in either PDF or PNG format.

-p PERCENT, --percent PERCENT
The maximum allowable percentage of missing CpGs. Set to 0.2 (20%) by default, which means that if more than 20% of the clock CpGs are missing, the estimation of DNAm age cannot be performed.

-d DELIMITER, --delimiter DELIMITER
Separator (usually TAB or comma) used in the input file. If the separator is not provided, the program will automatically detect the separator.

-f {pdf,png}, --format {pdf,png}
Figure format of the output coef plot. It must be "pdf" or "png". The default is "pdf".

-m meta_file, --metadata meta_file
This file contains the meta information for each sample. This file must have a header row, which contains the names or labels for variables (such as "Sex", "Age"). If the header row includes an "Age" column, a scatter plot will be generated to display the correlation between chronological age and predicted DNAm age. The first column of this file should contain sample IDs. The default value is None, indicating that no meta information file is provided.

-l mog_file, --log mog_file
This file is used to save the log information. By default, if no file is specified (None), the log information will be printed to the screen.

--impute {-1,0,1,2,3,4,5,6,7,8,9,10}
The imputation method code must be one of the 12 digits including (-1, 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10). The interpretations are: -1: Remove CpGs with any missing values. 0: Fill all missing values with '0.0'. 1: Fill all missing values with '1.0'. 2: Fill the missing values with **column mean** 3: Fill the missing values with **column median** 4: Fill the missing values with **column min** 5: Fill the missing values with **column max** 6: Fill the missing values with **row mean** 7: Fill the missing values with **row median** 8: Fill the missing values with **row min** 9: Fill the missing values with **row max** 10: Fill the missing values with **external reference** If 10 is specified, an external reference file must be

(continues on next page)

(continued from previous page)

```

        provided.
-r ref_file, --ref ref_file
    The external reference file contains two columns,
    separated by either tabs or commas. The first column
    represents the probe ID, while the second column
    contains the corresponding beta values.
--debug
    If set, print detailed information for debugging.
--overwrite
    If set, over-write existing output files.

```

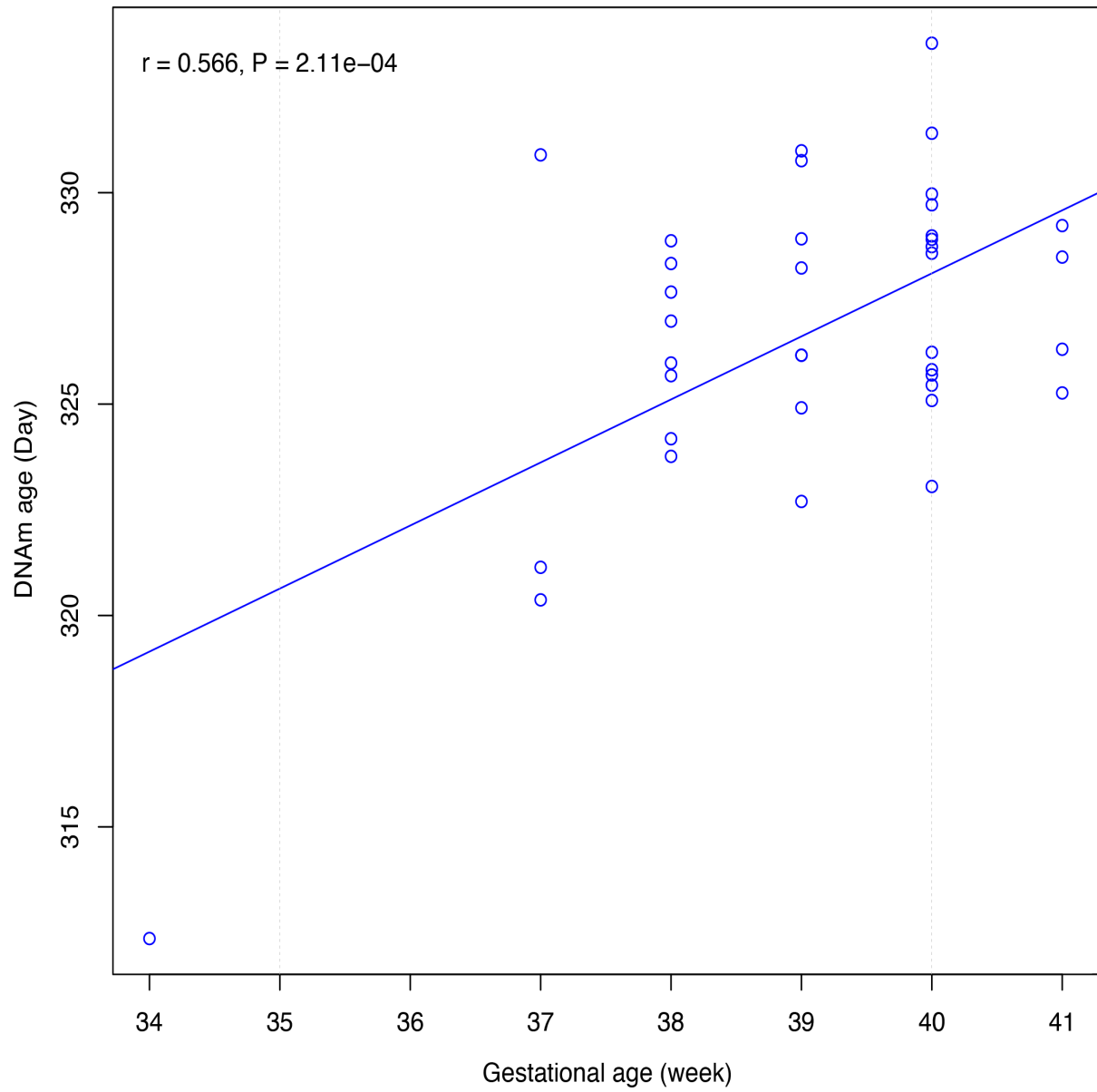
21.2 Example

```
$ epical GA_Haftorn Test5_CordBlood_N38_450K_beta.tsv.gz -m Test5_CordBlood_N38_450K_info.
tsv -o GA_Haftorn --overwrite -p 0.51
```

```

2024-01-23 10:56:02 [INFO] The prefix of output files is set to "GA_Haftorn".
2024-01-23 10:56:02 [WARNING] Over write existing files with prefix: GA_Haftorn
2024-01-23 10:56:02 [INFO] Loading GA_Haftorn clock data ...
2024-01-23 10:56:02 [INFO] Clock's name: "Haftorn_gestational"
2024-01-23 10:56:02 [INFO] Clock was trained from: "cord blood"
2024-01-23 10:56:02 [INFO] Clock's unit: "weeks"
2024-01-23 10:56:02 [INFO] Number of CpGs used: 176
...

```



GA_KNIGHT

Predictor CpGs	148
Unit	Week
Target Tissue	Neonatal Cord blood
Target Population	New born (Gestational age)
Method	Elastic Net regression
Reference	Knight AK, et al., Genome Biol. (2016).
Alias	

22.1 Usage

```
usage: epical GA_Knight [-h] [-o out_prefix] [-p PERCENT] [-d DELIMITER]
                        [-f {pdf,png}] [-m meta_file] [-l log_file]
                        [--impute {-1,0,1,2,3,4,5,6,7,8,9,10}] [-r ref_file]
                        [--debug] [--overwrite]
                        Input_file

positional arguments:
  Input_file            The input tabular structure file containing DNA
                        methylation data. This file must have a header row,
                        which contains the names or labels for samples. The
                        first column of this file should contain CpG IDs. The
                        remaining cells in the file should contain DNA
                        methylation beta values, represented as floating-
                        point numbers between 0 and 1. Use a TAB, comma, or any
                        other delimiter to separate the columns. Use 'NaN' or
                        'NA' to represent missing values. This file can be a
                        regular text file or compressed file (".gz", ".Z",
                        ".z", ".bz", ".bz2", ".bzip2").

options:
  -h, --help            show this help message and exit
  -o out_prefix, --output out_prefix
                        The PREFIX of output files. If no PREFIX is provided,
                        the default prefix "clock_name_out" is used. The
                        generated output files include:
                        "<PREFIX>.DNAm_age.tsv": This file contains the
                        predicted DNAm age. "<PREFIX>.used_CpGs.tsv": This
```

(continues on next page)

(continued from previous page)

file lists the CpGs that were used to calculate the DNAm age. "<PREFIX>.missed_CpGs.txt": This file provides a list of clock CpGs that were missed or excluded from the input file. "<PREFIX>.coef.tsv": This file contains a list of clock CpGs along with their coefficients. The last column indicates whether the CpG is included in the calculation.

"<PREFIX>.plots.R": This file is an R script used to generate visualization plots.

"<PREFIX>.coef_plot.pdf": This file is the coefficient plot in either PDF or PNG format.

-p PERCENT, --percent PERCENT
The maximum allowable percentage of missing CpGs. Set to 0.2 (20%) by default, which means that if more than 20% of the clock CpGs are missing, the estimation of DNAm age cannot be performed.

-d DELIMITER, --delimiter DELIMITER
Separator (usually TAB or comma) used in the input file. If the separator is not provided, the program will automatically detect the separator.

-f {pdf,png}, --format {pdf,png}
Figure format of the output coef plot. It must be "pdf" or "png". The default is "pdf".

-m meta_file, --metadata meta_file
This file contains the meta information for each sample. This file must have a header row, which contains the names or labels for variables (such as "Sex", "Age"). If the header row includes an "Age" column, a scatter plot will be generated to display the correlation between chronological age and predicted DNAm age. The first column of this file should contain sample IDs. The default value is None, indicating that no meta information file is provided.

-l log_file, --log log_file
This file is used to save the log information. By default, if no file is specified (None), the log information will be printed to the screen.

--impute {-1,0,1,2,3,4,5,6,7,8,9,10}
The imputation method code must be one of the 12 digits including (-1, 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10). The interpretations are: -1: Remove CpGs with any missing values. 0: Fill all missing values with '0.0'. 1: Fill all missing values with '1.0'. 2: Fill the missing values with ****column mean**** 3: Fill the missing values with ****column median**** 4: Fill the missing values with ****column min**** 5: Fill the missing values with ****column max**** 6: Fill the missing values with ****row mean**** 7: Fill the missing values with ****row median**** 8: Fill the missing values with ****row min**** 9: Fill the missing values with ****row max**** 10: Fill the missing values with ****external reference**** If 10 is specified, an external reference file must be

(continues on next page)

(continued from previous page)

```

        provided.
-r ref_file, --ref ref_file
    The external reference file contains two columns,
    separated by either tabs or commas. The first column
    represents the probe ID, while the second column
    contains the corresponding beta values.
--debug
    If set, print detailed information for debugging.
--overwrite
    If set, over-write existing output files.

```

22.2 Example

```

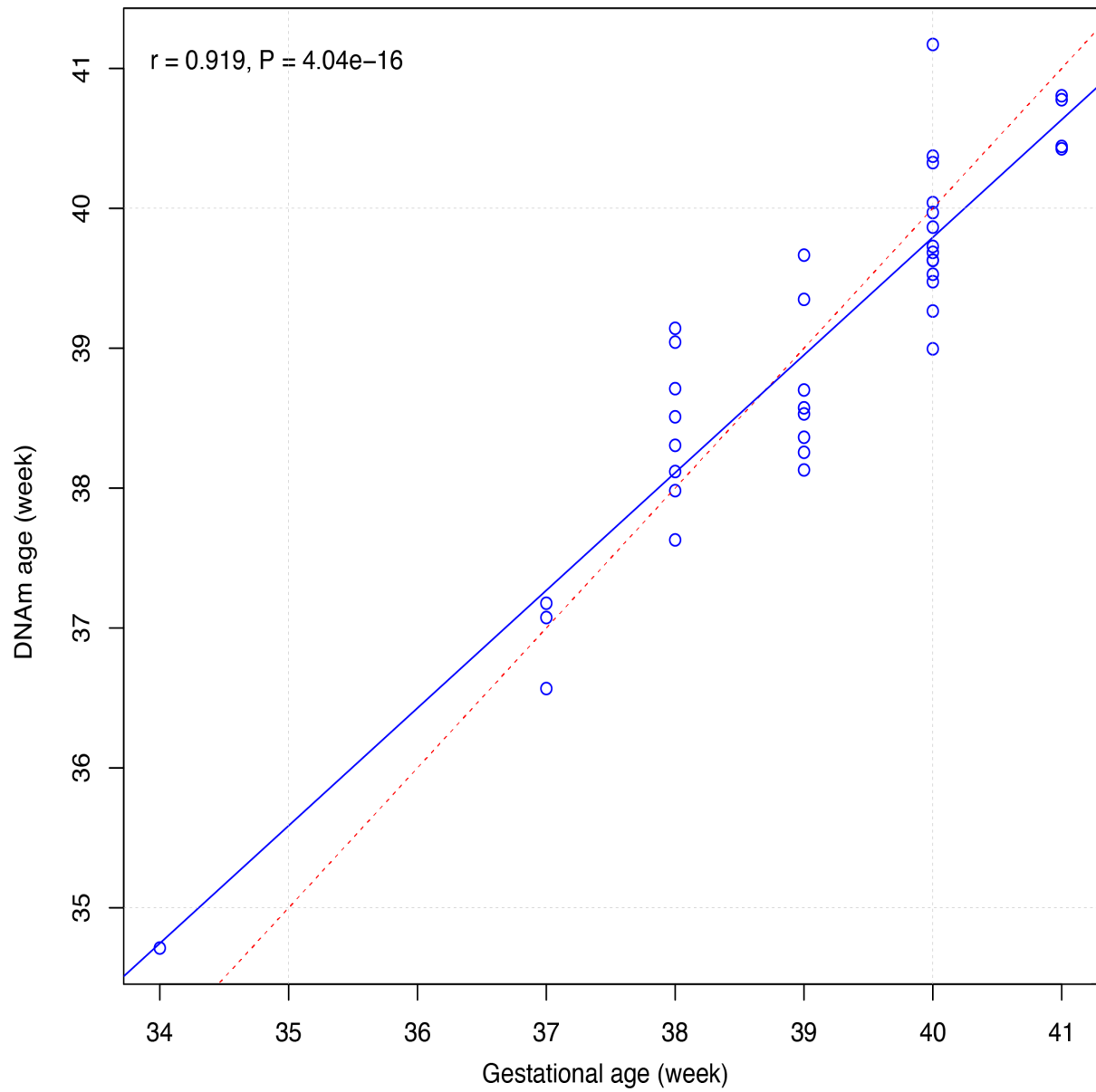
$ epical GA_Knight Test5_CordBlood_N38_450K_beta.tsv.gz -m Test5_CordBlood_N38_450K_info.
tsv -o GA_Knight --overwrite

```

```

2024-01-22 09:49:43 [INFO] The prefix of output files is set to "GA_Bohlin".
2024-01-22 09:49:43 [WARNING] Over write existing files with prefix: GA_Bohlin
2024-01-22 09:49:43 [INFO] Loading GA_Bohlin clock data ...
2024-01-22 09:49:43 [INFO] Clock's name: "Bohlin_gestational"
2024-01-22 09:49:43 [INFO] Clock was trained from: "cord blood"
2024-01-22 09:49:43 [INFO] Clock's unit: "days"
2024-01-22 09:49:43 [INFO] Number of CpGs used: 96
...

```



GA_MAYNE

Predictor CpGs	62
Unit	Week
Target Tissue	Placental
Target Population	New born (Gestational age)
Method	Elastic Net regression
Reference	Mayne BT, et al, Epigenomics. (2017).
Alias	

23.1 Usage

```
usage: epical GA_Mayne [-h] [-o out_prefix] [-p PERCENT] [-d DELIMITER]
                        [-f {pdf,png}] [-m meta_file] [-l log_file]
                        [--impute {-1,0,1,2,3,4,5,6,7,8,9,10}] [-r ref_file]
                        [--debug] [--overwrite]
                        Input_file

positional arguments:
  Input_file            The input tabular structure file containing DNA
                        methylation data. This file must have a header row,
                        which contains the names or labels for samples
                        The first column of this file should contain CpG IDs.
                        The remaining cells in the file should contain DNA
                        methylation beta values, represented as floating-
                        point numbers between 0 and 1. Use a TAB, comma, or any
                        other delimiter to separate the columns. Use 'NaN' or
                        'NA' to represent missing values. This file can be a
                        regular text file or compressed file (".gz", ".Z",
                        ".z", ".bz", ".bz2", ".bzip2").

options:
  -h, --help            show this help message and exit
  -o out_prefix, --output out_prefix
                        The PREFIX of output files. If no PREFIX is provided,
                        the default prefix "clock_name_out" is used. The
                        generated output files include:
                        "<PREFIX>.DNAm_age.tsv": This file contains the
                        predicted DNAm age. "<PREFIX>.used_CpGs.tsv": This
```

(continues on next page)

(continued from previous page)

file lists the CpGs that were used to calculate the DNAm age. "<PREFIX>.missed_CpGs.txt": This file provides a list of clock CpGs that were missed or excluded from the input file. "<PREFIX>.coef.tsv": This file contains a list of clock CpGs along with their coefficients. The last column indicates whether the CpG is included in the calculation.

"<PREFIX>.plots.R": This file is an R script used to generate visualization plots.

"<PREFIX>.coef_plot.pdf": This file is the coefficient plot in either PDF or PNG format.

-p PERCENT, --percent PERCENT
The maximum allowable percentage of missing CpGs. Set to 0.2 (20%) by default, which means that if more than 20% of the clock CpGs are missing, the estimation of DNAm age cannot be performed.

-d DELIMITER, --delimiter DELIMITER
Separator (usually TAB or comma) used in the input file. If the separator is not provided, the program will automatically detect the separator.

-f {pdf,png}, --format {pdf,png}
Figure format of the output coef plot. It must be "pdf" or "png". The default is "pdf".

-m meta_file, --metadata meta_file
This file contains the meta information for each sample. This file must have a header row, which contains the names or labels for variables (such as "Sex", "Age"). If the header row includes an "Age" column, a scatter plot will be generated to display the correlation between chronological age and predicted DNAm age. The first column of this file should contain sample IDs. The default value is None, indicating that no meta information file is provided.

-l log_file, --log log_file
This file is used to save the log information. By default, if no file is specified (None), the log information will be printed to the screen.

--impute {-1,0,1,2,3,4,5,6,7,8,9,10}
The imputation method code must be one of the 12 digits including (-1, 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10). The interpretations are: -1: Remove CpGs with any missing values. 0: Fill all missing values with '0.0'. 1: Fill all missing values with '1.0'. 2: Fill the missing values with **column mean** 3: Fill the missing values with **column median** 4: Fill the missing values with **column min** 5: Fill the missing values with **column max** 6: Fill the missing values with **row mean** 7: Fill the missing values with **row median** 8: Fill the missing values with **row min** 9: Fill the missing values with **row max** 10: Fill the missing values with **external reference** If 10 is specified, an external reference file must be

(continues on next page)

(continued from previous page)

```

        provided.
-r ref_file, --ref ref_file
    The external reference file contains two columns,
    separated by either tabs or commas. The first column
    represents the probe ID, while the second column
    contains the corresponding beta values.
--debug
    If set, print detailed information for debugging.
--overwrite
    If set, over-write existing output files.

```

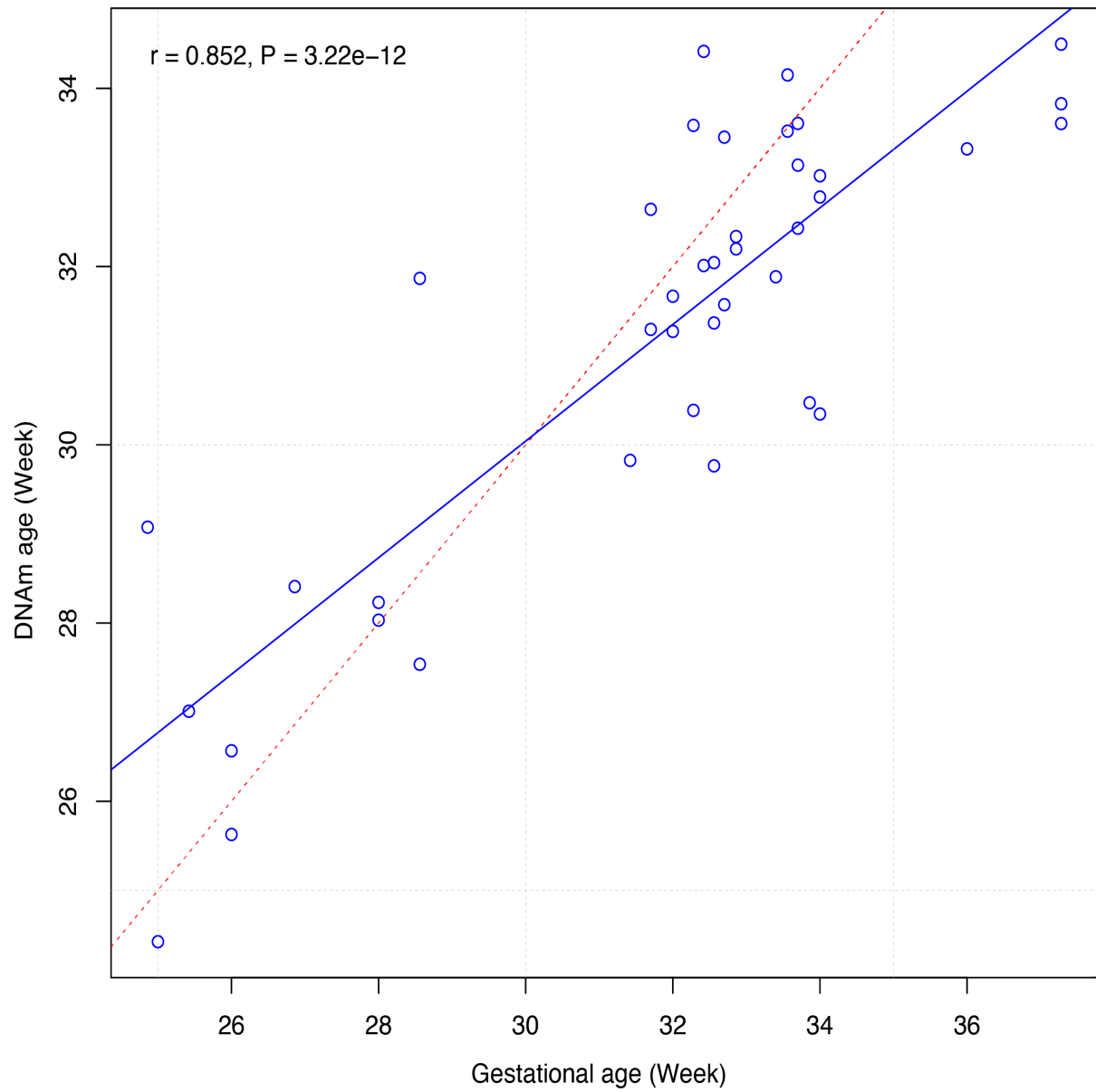
23.2 Example

```
$ epical GA_Mayne Test6_Placentas_N40_450K_beta.tsv.gz -m Test6_Placentas_N40_450K_info.
tsv -o GA_Mayne --overwrite
```

```

2024-01-23 11:44:31 [INFO] The prefix of output files is set to "GA_Mayne".
2024-01-23 11:44:31 [WARNING] Over write existing files with prefix: GA_Mayne
2024-01-23 11:44:31 [INFO] Loading GA_Mayne clock data ...
2024-01-23 11:44:31 [INFO] Clock's name: "Mayne_gestational"
2024-01-23 11:44:31 [INFO] Clock was trained from: "placental"
2024-01-23 11:44:31 [INFO] Clock's unit: "weeks"
2024-01-23 11:44:31 [INFO] Number of CpGs used: 62
...

```



GA_LEE_CPC

Predictor CpGs	1125
Unit	Week
Target Tissue	Placental
Target Population	New born (Gestational age)
Method	Elastic Net regression
Reference	Lee Y, et al, Aging. (2019).
Alias	A control placental clock (CPC)

24.1 Usage

```
usage: epical GA_Lee_CPC [-h] [-o out_prefix] [-p PERCENT] [-d DELIMITER]
                        [-f {pdf,png}] [-m meta_file] [-l log_file]
                        [--impute {-1,0,1,2,3,4,5,6,7,8,9,10}] [-r ref_file]
                        [--debug] [--overwrite]
                        Input_file

positional arguments:
  Input_file            The input tabular structure file containing DNA
                        methylation data. This file must have a header row,
                        which contains the names or labels for samples. The
                        first column of this file should contain CpG IDs. The
                        remaining cells in the file should contain DNA
                        methylation beta values, represented as floating-
                        point numbers between 0 and 1. Use a TAB, comma, or any
                        other delimiter to separate the columns. Use 'NaN' or
                        'NA' to represent missing values. This file can be a
                        regular text file or compressed file (".gz", ".Z",
                        ".z", ".bz", ".bz2", ".bzip2").

options:
  -h, --help            show this help message and exit
  -o out_prefix, --output out_prefix
                        The PREFIX of output files. If no PREFIX is provided,
                        the default prefix "clock_name_out" is used. The
                        generated output files include:
                        "<PREFIX>.DNAm_age.tsv": This file contains the
                        predicted DNAm age. "<PREFIX>.used_CpGs.tsv": This
```

(continues on next page)

(continued from previous page)

file lists the CpGs that were used to calculate the DNAm age. "<PREFIX>.missed_CpGs.txt": This file provides a list of clock CpGs that were missed or excluded from the input file. "<PREFIX>.coef.tsv": This file contains a list of clock CpGs along with their coefficients. The last column indicates whether the CpG is included in the calculation.

"<PREFIX>.plots.R": This file is an R script used to generate visualization plots.

"<PREFIX>.coef_plot.pdf": This file is the coefficient plot in either PDF or PNG format.

-p PERCENT, --percent PERCENT
The maximum allowable percentage of missing CpGs. Set to 0.2 (20%) by default, which means that if more than 20% of the clock CpGs are missing, the estimation of DNAm age cannot be performed.

-d DELIMITER, --delimiter DELIMITER
Separator (usually TAB or comma) used in the input file. If the separator is not provided, the program will automatically detect the separator.

-f {pdf,png}, --format {pdf,png}
Figure format of the output coef plot. It must be "pdf" or "png". The default is "pdf".

-m meta_file, --metadata meta_file
This file contains the meta information for each sample. This file must have a header row, which contains the names or labels for variables (such as "Sex", "Age"). If the header row includes an "Age" column, a scatter plot will be generated to display the correlation between chronological age and predicted DNAm age. The first column of this file should contain sample IDs. The default value is None, indicating that no meta information file is provided.

-l log_file, --log log_file
This file is used to save the log information. By default, if no file is specified (None), the log information will be printed to the screen.

--impute {-1,0,1,2,3,4,5,6,7,8,9,10}
The imputation method code must be one of the 12 digits including (-1, 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10). The interpretations are: -1: Remove CpGs with any missing values. 0: Fill all missing values with '0.0'. 1: Fill all missing values with '1.0'. 2: Fill the missing values with **column mean** 3: Fill the missing values with **column median** 4: Fill the missing values with **column min** 5: Fill the missing values with **column max** 6: Fill the missing values with **row mean** 7: Fill the missing values with **row median** 8: Fill the missing values with **row min** 9: Fill the missing values with **row max** 10: Fill the missing values with **external reference** If 10 is specified, an external reference file must be

(continues on next page)

(continued from previous page)

```

        provided.
-r ref_file, --ref ref_file
    The external reference file contains two columns,
    separated by either tabs or commas. The first column
    represents the probe ID, while the second column
    contains the corresponding beta values.
--debug
    If set, print detailed information for debugging.
--overwrite
    If set, over-write existing output files.

```

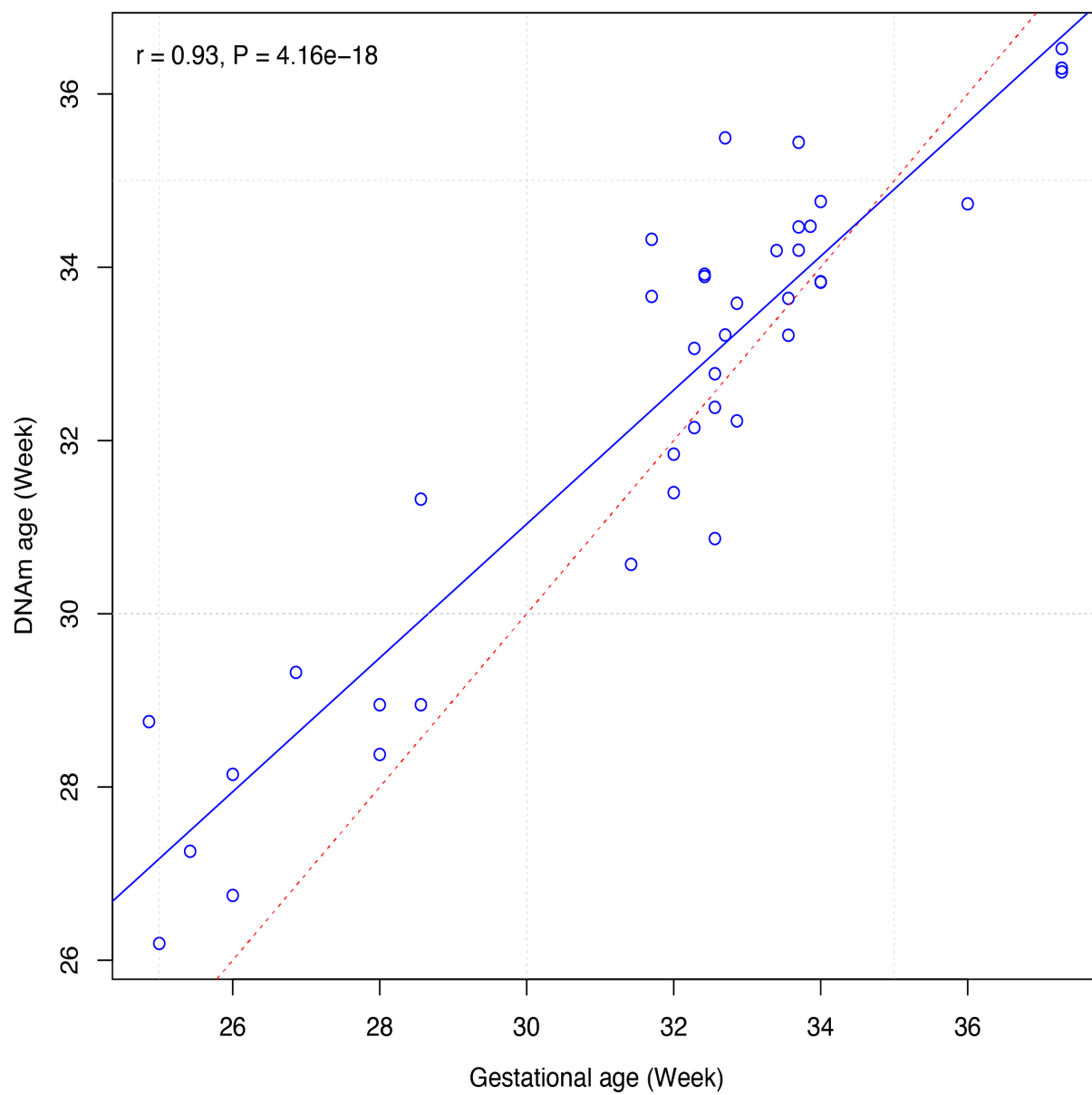
24.2 Example

```
$ epical GA_Lee_CPC Test6_Placentas_N40_450K_beta.tsv.gz -m Test6_Placentas_N40_450K_info.
tsv -o GA_Lee_CPC
```

```

2024-01-23 12:07:03 [INFO] The prefix of output files is set to "GA_Lee_CPC".
2024-01-23 12:07:03 [INFO] Loading GA_Lee_CPC clock data ...
2024-01-23 12:07:03 [INFO] Clock's name: "Lee_gestational_CPC"
2024-01-23 12:07:03 [INFO] Clock was trained from: "placental"
2024-01-23 12:07:03 [INFO] Clock's unit: "weeks"
2024-01-23 12:07:03 [INFO] Number of CpGs used: 1125
...

```



GA_LEE_RPC

Predictor CpGs	1125
Unit	Week
Target Tissue	Placental
Target Population	New born (Gestational age)
Method	Elastic Net regression
Reference	Lee Y, et al, Aging. (2019).
Alias	A robust placental clock (RPC)

25.1 Usage

```
usage: epical GA_Lee_RPC [-h] [-o out_prefix] [-p PERCENT] [-d DELIMITER]
                        [-f {pdf,png}] [-m meta_file] [-l log_file]
                        [--impute {-1,0,1,2,3,4,5,6,7,8,9,10}] [-r ref_file]
                        [--debug] [--overwrite]
                        Input_file

positional arguments:
  Input_file            The input tabular structure file containing DNA
                        methylation data. This file must have a header row,
                        which contains the names or labels for samples. The
                        first column of this file should contain CpG IDs. The
                        remaining cells in the file should contain DNA
                        methylation beta values, represented as floating-
                        point numbers between 0 and 1. Use a TAB, comma, or any
                        other delimiter to separate the columns. Use 'NaN' or
                        'NA' to represent missing values. This file can be a
                        regular text file or compressed file (".gz", ".Z",
                        ".z", ".bz", ".bz2", ".bzip2").

options:
  -h, --help            show this help message and exit
  -o out_prefix, --output out_prefix
                        The PREFIX of output files. If no PREFIX is provided,
                        the default prefix "clock_name_out" is used. The
                        generated output files include:
                        "<PREFIX>.DNAm_age.tsv": This file contains the
                        predicted DNAm age. "<PREFIX>.used_CpGs.tsv": This
```

(continues on next page)

(continued from previous page)

file lists the CpGs that were used to calculate the DNAm age. "<PREFIX>.missed_CpGs.txt": This file provides a list of clock CpGs that were missed or excluded from the input file. "<PREFIX>.coef.tsv": This file contains a list of clock CpGs along with their coefficients. The last column indicates whether the CpG is included in the calculation.

"<PREFIX>.plots.R": This file is an R script used to generate visualization plots.

"<PREFIX>.coef_plot.pdf": This file is the coefficient plot in either PDF or PNG format.

-p PERCENT, --percent PERCENT
The maximum allowable percentage of missing CpGs. Set to 0.2 (20%) by default, which means that if more than 20% of the clock CpGs are missing, the estimation of DNAm age cannot be performed.

-d DELIMITER, --delimiter DELIMITER
Separator (usually TAB or comma) used in the input file. If the separator is not provided, the program will automatically detect the separator.

-f {pdf,png}, --format {pdf,png}
Figure format of the output coef plot. It must be "pdf" or "png". The default is "pdf".

-m meta_file, --metadata meta_file
This file contains the meta information for each sample. This file must have a header row, which contains the names or labels for variables (such as "Sex", "Age"). If the header row includes an "Age" column, a scatter plot will be generated to display the correlation between chronological age and predicted DNAm age. The first column of this file should contain sample IDs. The default value is None, indicating that no meta information file is provided.

-l log_file, --log log_file
This file is used to save the log information. By default, if no file is specified (None), the log information will be printed to the screen.

--impute {-1,0,1,2,3,4,5,6,7,8,9,10}
The imputation method code must be one of the 12 digits including (-1, 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10). The interpretations are: -1: Remove CpGs with any missing values. 0: Fill all missing values with '0.0'. 1: Fill all missing values with '1.0'. 2: Fill the missing values with **column mean** 3: Fill the missing values with **column median** 4: Fill the missing values with **column min** 5: Fill the missing values with **column max** 6: Fill the missing values with **row mean** 7: Fill the missing values with **row median** 8: Fill the missing values with **row min** 9: Fill the missing values with **row max** 10: Fill the missing values with **external reference** If 10 is specified, an external reference file must be

(continues on next page)

(continued from previous page)

```

        provided.
-r ref_file, --ref ref_file
    The external reference file contains two columns,
    separated by either tabs or commas. The first column
    represents the probe ID, while the second column
    contains the corresponding beta values.
--debug
    If set, print detailed information for debugging.
--overwrite
    If set, over-write existing output files.

```

25.2 Example

```

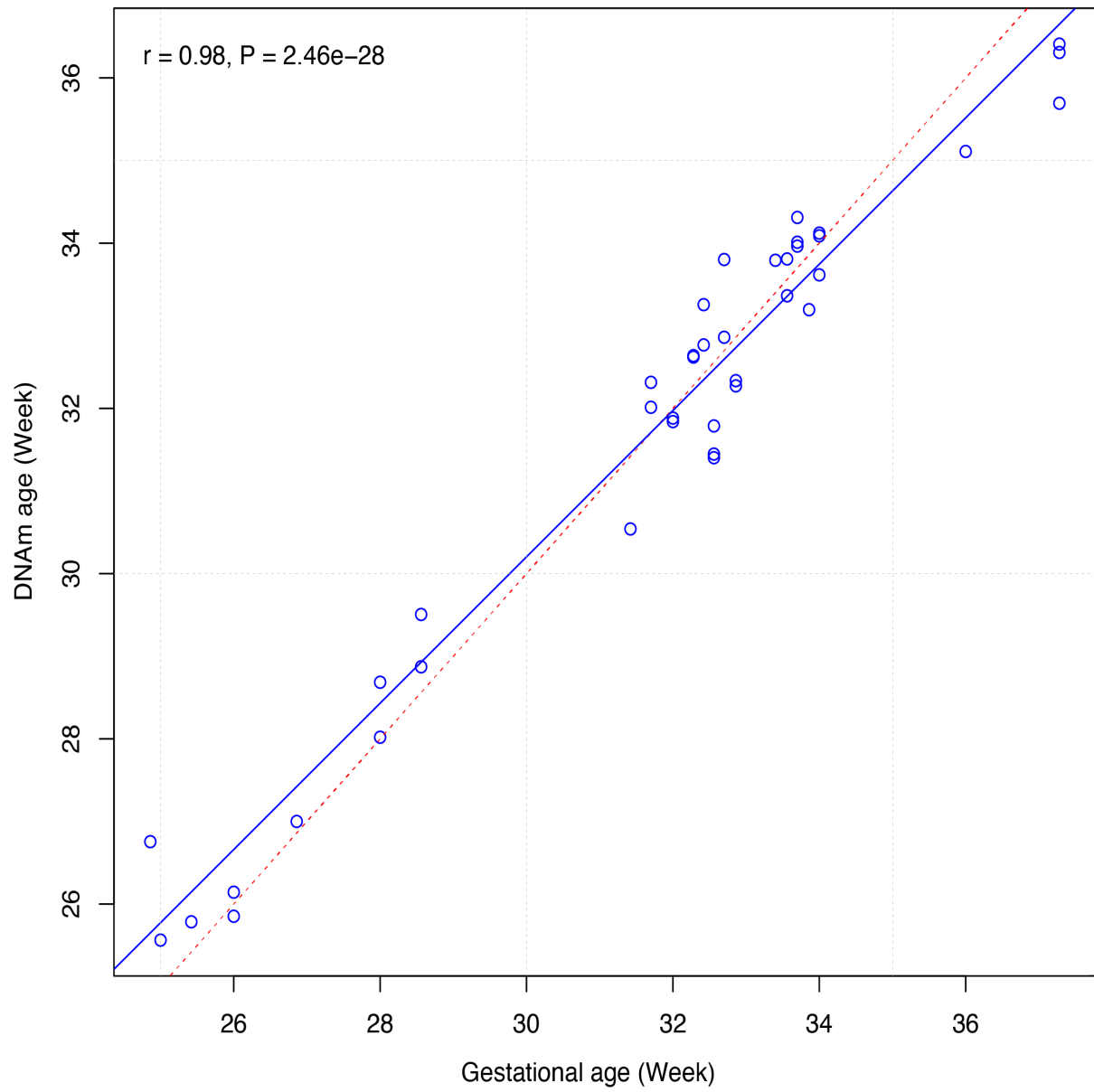
$ epical GA_Lee_RPC Test6_Placentas_N40_450K_beta.tsv.gz -m Test6_Placentas_N40_450K_info.
tsv -o GA_Lee_RPC

```

```

2024-01-23 12:07:03 [INFO] The prefix of output files is set to "GA_Lee_RPC".
2024-01-23 12:07:03 [INFO] Loading GA_Lee_CPC clock data ...
2024-01-23 12:07:03 [INFO] Clock's name: "Lee_gestational_RPC"
2024-01-23 12:07:03 [INFO] Clock was trained from: "placental"
2024-01-23 12:07:03 [INFO] Clock's unit: "weeks"
2024-01-23 12:07:03 [INFO] Number of CpGs used: 1125
...

```



GA_LEE_RRPC

Predictor CpGs	1125
Unit	Week
Target Tissue	Placental
Target Population	New born (Gestational age)
Method	Elastic Net regression
Reference	Lee Y, et al, Aging. (2019).
Alias	refined robust placental clock, rRPC

26.1 Usage

```
usage: epical GA_Lee_rRPC [-h] [-o out_prefix] [-p PERCENT] [-d DELIMITER]
                        [-f {pdf,png}] [-m meta_file] [-l log_file]
                        [--impute {-1,0,1,2,3,4,5,6,7,8,9,10}] [-r ref_file]
                        [--debug] [--overwrite]
                        Input_file
```

positional arguments:

Input_file The input tabular structure file containing DNA methylation data. This file must have a header row, which contains the names or labels for samples. The first column of this file should contain CpG IDs. The remaining cells in the file should contain DNA methylation beta values, represented as floating-point numbers between 0 and 1. Use a TAB, comma, or any other delimiter to separate the columns. Use 'NaN' or 'NA' to represent missing values. This file can be a regular text file or compressed file (".gz", ".Z", ".z", ".bz", ".bz2", ".bzip2").

options:

-h, --help show this help message and exit

-o out_prefix, --output out_prefix The PREFIX of output files. If no PREFIX is provided, the default prefix "clock_name_out" is used. The generated output files include:
 "<PREFIX>.DNAm_age.tsv": This file contains the predicted DNAm age. "<PREFIX>.used_CpGs.tsv": This

(continues on next page)

(continued from previous page)

file lists the CpGs that were used to calculate the DNAm age. "<PREFIX>.missed_CpGs.txt": This file provides a list of clock CpGs that were missed or excluded from the input file. "<PREFIX>.coef.tsv": This file contains a list of clock CpGs along with their coefficients. The last column indicates whether the CpG is included in the calculation.

"<PREFIX>.plots.R": This file is an R script used to generate visualization plots.

"<PREFIX>.coef_plot.pdf": This file is the coefficient plot in either PDF or PNG format.

-p PERCENT, --percent PERCENT
The maximum allowable percentage of missing CpGs. Set to 0.2 (20%) by default, which means that if more than 20% of the clock CpGs are missing, the estimation of DNAm age cannot be performed.

-d DELIMITER, --delimiter DELIMITER
Separator (usually TAB or comma) used in the input file. If the separator is not provided, the program will automatically detect the separator.

-f {pdf,png}, --format {pdf,png}
Figure format of the output coef plot. It must be "pdf" or "png". The default is "pdf".

-m meta_file, --metadata meta_file
This file contains the meta information for each sample. This file must have a header row, which contains the names or labels for variables (such as "Sex", "Age"). If the header row includes an "Age" column, a scatter plot will be generated to display the correlation between chronological age and predicted DNAm age. The first column of this file should contain sample IDs. The default value is None, indicating that no meta information file is provided.

-l log_file, --log log_file
This file is used to save the log information. By default, if no file is specified (None), the log information will be printed to the screen.

--impute {-1,0,1,2,3,4,5,6,7,8,9,10}
The imputation method code must be one of the 12 digits including (-1, 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10). The interpretations are: -1: Remove CpGs with any missing values. 0: Fill all missing values with '0.0'. 1: Fill all missing values with '1.0'. 2: Fill the missing values with ****column mean**** 3: Fill the missing values with ****column median**** 4: Fill the missing values with ****column min**** 5: Fill the missing values with ****column max**** 6: Fill the missing values with ****row mean**** 7: Fill the missing values with ****row median**** 8: Fill the missing values with ****row min**** 9: Fill the missing values with ****row max**** 10: Fill the missing values with ****external reference**** If 10 is specified, an external reference file must be

(continues on next page)

(continued from previous page)

```

        provided.
-r ref_file, --ref ref_file
    The external reference file contains two columns,
    separated by either tabs or commas. The first column
    represents the probe ID, while the second column
    contains the corresponding beta values.
--debug
    If set, print detailed information for debugging.
--overwrite
    If set, over-write existing output files.

```

26.2 Example

```

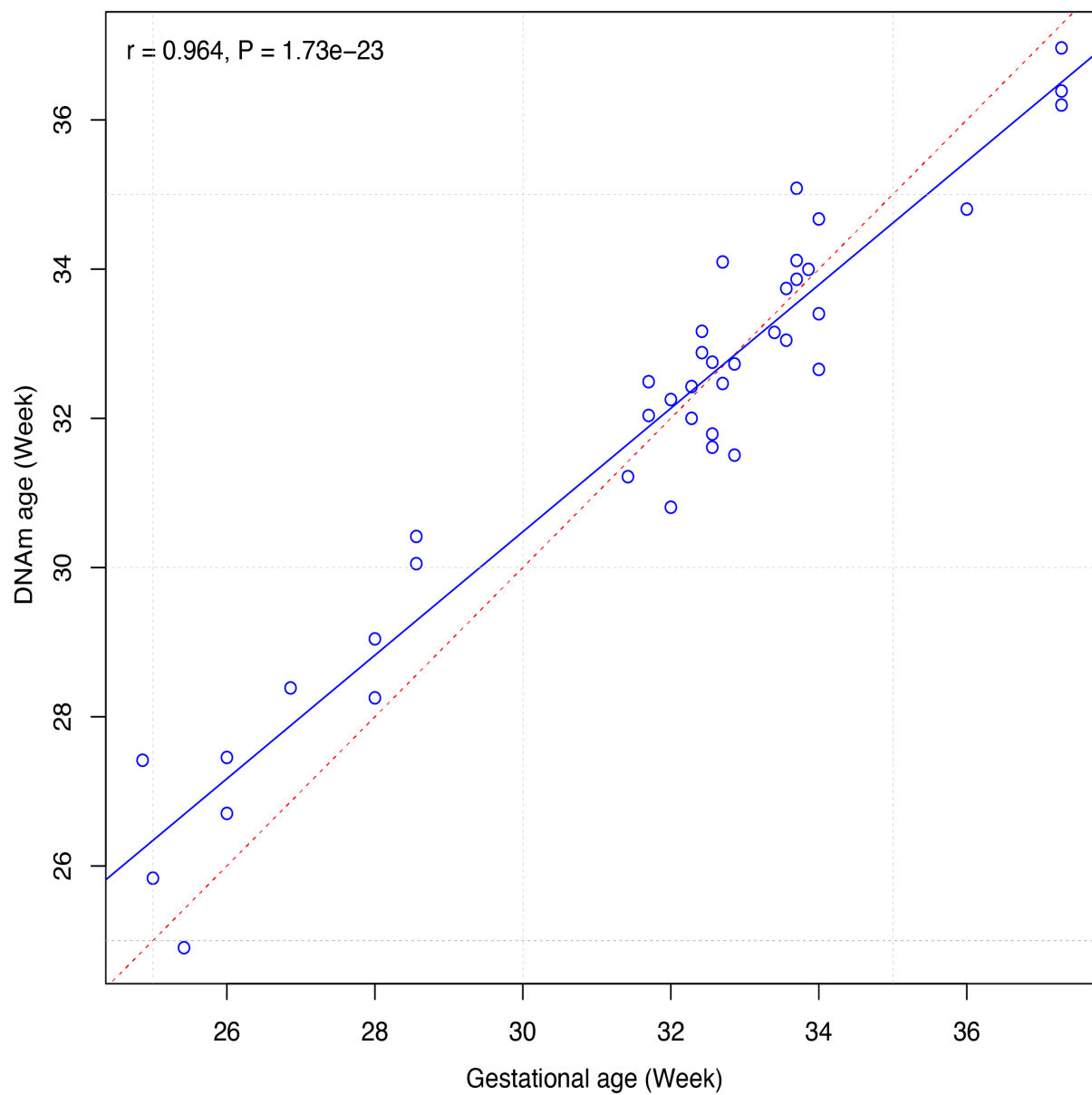
$ epical GA_Lee_rRPC Test6_Placentas_N40_450K_beta.tsv.gz -m Test6_Placentas_N40_450K_info.
tsv -o GA_Lee_rRPC

```

```

2024-01-23 12:07:03 [INFO] The prefix of output files is set to "GA_Lee_RPC".
2024-01-23 12:07:03 [INFO] Loading GA_Lee_CPC clock data ...
2024-01-23 12:07:03 [INFO] Clock's name: "Lee_gestational_RPC"
2024-01-23 12:07:03 [INFO] Clock was trained from: "placental"
2024-01-23 12:07:03 [INFO] Clock's unit: "weeks"
2024-01-23 12:07:03 [INFO] Number of CpGs used: 1125
...

```



Predictor CpGs	435
Unit	Day
Target Tissue	Multi-tissue
Target Population	Mouse (whole life)
Method	Elastic Net regression
Reference	Meer MV, et al, Elife. (2018).
Alias	Mouse whole life multiple tissue clock

27.1 Usage

```
usage: epical WLMT [-h] [-o out_prefix] [-g {mm10,mm39}] [-p PERCENT]
                  [-d DELIMITER] [-f {pdf,png}] [-m meta_file] [-l log_file]
                  [--impute {-1,0,1,2,3,4,5,6,7,8,9,10}] [-r ref_file]
                  [--debug] [--overwrite]
                  Input_file
```

positional arguments:

Input_file The input tabular structure file containing DNA methylation data. This file must have a header row, which contains the names or labels for samples. The first column of this file should contain CpG IDs. The remaining cells in the file should contain DNA methylation beta values, represented as floating-point numbers between 0 and 1. Use a TAB, comma, or any other delimiter to separate the columns. Use 'NaN' or 'NA' to represent missing values. This file can be a regular text file or compressed file (".gz", ".Z", ".z", ".bz", ".bz2", ".bzip2").

options:

-h, --help show this help message and exit

-o out_prefix, --output out_prefix The PREFIX of output files. If no PREFIX is provided, the default prefix "clock_name_out" is used. The generated output files include:
 "<PREFIX>.DNAm_age.tsv": This file contains the predicted DNAm age. "<PREFIX>.used_CpGs.tsv": This

(continues on next page)

(continued from previous page)

file lists the CpGs that were used to calculate the DNAm age. "<PREFIX>.missed_CpGs.txt": This file provides a list of clock CpGs that were missed or excluded from the input file. "<PREFIX>.coef.tsv": This file contains a list of clock CpGs along with their coefficients. The last column indicates whether the CpG is included in the calculation.

"<PREFIX>.plots.R": This file is an R script used to generate visualization plots.

"<PREFIX>.coef_plot.pdf": This file is the coefficient plot in either PDF or PNG format.

-g {mm10,mm39}, --genome {mm10,mm39}
The reference genome for Mouse (*Mus musculus*) used for RRBS or WGBS reads alignment. Must be 'mm10' or 'mm39'. default is 'mm10'.

-p PERCENT, --percent PERCENT
The maximum allowable percentage of missing CpGs. Set to 0.2 (20%) by default, which means that if more than 20% of the clock CpGs are missing, the estimation of DNAm age cannot be performed.

-d DELIMITER, --delimiter DELIMITER
Separator (usually TAB or comma) used in the input file. If the separator is not provided, the program will automatically detect the separator.

-f {pdf,png}, --format {pdf,png}
The output plot format must be either 'pdf' or 'png'. The default format is 'pdf'."

-m meta_file, --metadata meta_file
This file contains the meta information for each sample. This file must have a header row, which contains the names or labels for variables (such as "Sex", "Age"). If the header row includes an "Age" column, a scatter plot will be generated to display the correlation between chronological age and predicted DNAm age. The first column of this file should contain sample IDs. The default value is None, indicating that no meta information file is provided.

-l log_file, --log log_file
This file is used to save the log information. By default, if no file is specified (None), the log information will be printed to the screen.

--impute {-1,0,1,2,3,4,5,6,7,8,9,10}
The imputation method code must be one of the 12 digits including (-1, 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10). The interpretations are: -1: Remove CpGs with any missing values. 0: Fill all missing values with '0.0'. 1: Fill all missing values with '1.0'. 2: Fill the missing values with ****column mean**** 3: Fill the missing values with ****column median**** 4: Fill the missing values with ****column min**** 5: Fill the missing values with ****column max**** 6: Fill the missing values with ****row mean**** 7: Fill the missing values with

(continues on next page)

(continued from previous page)

```

**row median** 8: Fill the missing values with **row
min** 9: Fill the missing values with **row max** 10:
Fill the missing values with **external reference** If
10 is specified, an external reference file must be
provided.
-r ref_file, --ref ref_file
    The external reference file contains two columns,
    separated by either tabs or commas. The first column
    represents the probe ID, while the second column
    contains the corresponding beta values.
--debug
    If set, print detailed information for debugging.
--overwrite
    If set, over-write existing output files.

```

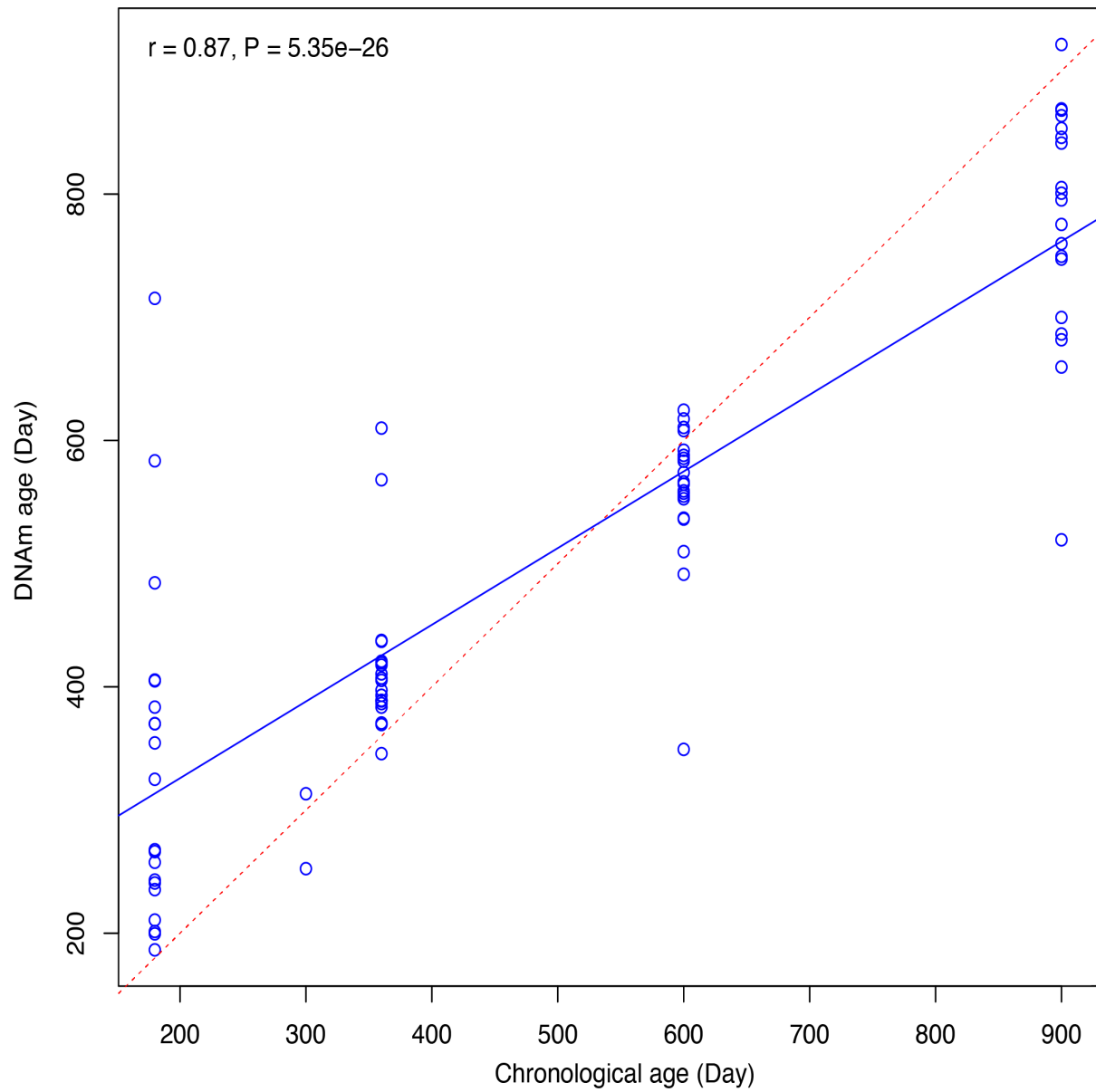
27.2 Example

```
$ epical WLMT Test7_mouse_N81_RRBS_beta.tsv -m Test7_mouse_N81_RRBS_info.tsv -g mm10 -o WLMT_out
```

```

2024-04-11 01:47:13 [INFO] The prefix of output files is set to "WLMT_out".
2024-04-11 01:47:13 [INFO] Loading WLMT clock data ...
2024-04-11 01:47:13 [INFO] Clock's name: "WLMT_mm10"
2024-04-11 01:47:13 [INFO] Clock was trained from: "Pan-tissue"
2024-04-11 01:47:13 [INFO] Clock's unit: "week"
2024-04-11 01:47:13 [INFO] Number of CpGs used: 435
...

```



Predictor CpGs	329
Unit	Day
Target Tissue	Multi-tissue
Target Population	Mouse (young age up to 10 month)
Method	Elastic Net regression
Reference	Stubbs TM, Genome biology (2017) .
Alias	

28.1 Usage

```
usage: epical YOMT [-h] [-o out_prefix] [-g {mm10,mm39}] [-p PERCENT]
                  [-d DELIMITER] [-f {pdf,png}] [-m meta_file] [-l log_file]
                  [--impute {-1,0,1,2,3,4,5,6,7,8,9,10}] [-r ref_file]
                  [--debug] [--overwrite]
                  Input_file
```

positional arguments:

Input_file The input tabular structure file containing DNA methylation data. This file must have a header row, which contains the names or labels for samples. The first column of this file should contain CpG IDs. The remaining cells in the file should contain DNA methylation beta values, represented as floating-point numbers between 0 and 1. Use a TAB, comma, or any other delimiter to separate the columns. Use 'NaN' or 'NA' to represent missing values. This file can be a regular text file or compressed file (".gz", ".Z", ".z", ".bz", ".bz2", ".bzip2").

options:

-h, --help show this help message and exit

-o out_prefix, --output out_prefix The PREFIX of output files. If no PREFIX is provided, the default prefix "clock_name_out" is used. The generated output files include:
 "<PREFIX>.DNAm_age.tsv": This file contains the predicted DNAm age. "<PREFIX>.used_CpGs.tsv": This

(continues on next page)

(continued from previous page)

file lists the CpGs that were used to calculate the DNAm age. "<PREFIX>.missed_CpGs.txt": This file provides a list of clock CpGs that were missed or excluded from the input file. "<PREFIX>.coef.tsv": This file contains a list of clock CpGs along with their coefficients. The last column indicates whether the CpG is included in the calculation.

"<PREFIX>.plots.R": This file is an R script used to generate visualization plots.

"<PREFIX>.coef_plot.pdf": This file is the coefficient plot in either PDF or PNG format.

-g {mm10,mm39}, --genome {mm10,mm39}
The reference genome for Mouse (*Mus musculus*) used for RRBS or WGBS reads alignment. Must be 'mm10' or 'mm39'. default is 'mm10'.

-p PERCENT, --percent PERCENT
The maximum allowable percentage of missing CpGs. Set to 0.2 (20%) by default, which means that if more than 20% of the clock CpGs are missing, the estimation of DNAm age cannot be performed.

-d DELIMITER, --delimiter DELIMITER
Separator (usually TAB or comma) used in the input file. If the separator is not provided, the program will automatically detect the separator.

-f {pdf,png}, --format {pdf,png}
The output plot format must be either 'pdf' or 'png'. The default format is 'pdf'."

-m meta_file, --metadata meta_file
This file contains the meta information for each sample. This file must have a header row, which contains the names or labels for variables (such as "Sex", "Age"). If the header row includes an "Age" column, a scatter plot will be generated to display the correlation between chronological age and predicted DNAm age. The first column of this file should contain sample IDs. The default value is None, indicating that no meta information file is provided.

-l log_file, --log log_file
This file is used to save the log information. By default, if no file is specified (None), the log information will be printed to the screen.

--impute {-1,0,1,2,3,4,5,6,7,8,9,10}
The imputation method code must be one of the 12 digits including (-1, 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10). The interpretations are: -1: Remove CpGs with any missing values. 0: Fill all missing values with '0.0'. 1: Fill all missing values with '1.0'. 2: Fill the missing values with ****column mean**** 3: Fill the missing values with ****column median**** 4: Fill the missing values with ****column min**** 5: Fill the missing values with ****column max**** 6: Fill the missing values with ****row mean**** 7: Fill the missing values with

(continues on next page)

(continued from previous page)

```

**row median** 8: Fill the missing values with **row
min** 9: Fill the missing values with **row max** 10:
Fill the missing values with **external reference** If
10 is specified, an external reference file must be
provided.
-r ref_file, --ref ref_file
    The external reference file contains two columns,
    separated by either tabs or commas. The first column
    represents the probe ID, while the second column
    contains the corresponding beta values.
--debug
    If set, print detailed information for debugging.
--overwrite
    If set, over-write existing output files.

```

28.2 Example

```
$ epical YOMT Test7_mouse_N81_RRBS_beta.tsv -g mm10 -o YOMT_out
```

```

2024-04-12 11:26:33 [INFO] The prefix of output files is set to "YOMT_out".
2024-04-12 11:26:33 [WARNING] Over write existing files with prefix: YOMT_out
2024-04-12 11:26:33 [INFO] Loading YOMT clock data ...
2024-04-12 11:26:33 [INFO] Clock's name: "YOMT_mm10"
2024-04-12 11:26:33 [INFO] Clock was trained from: "Multi-tissue"
2024-04-12 11:26:33 [INFO] Clock's unit: "week"
2024-04-12 11:26:33 [INFO] Number of CpGs used: 329
...

```


MMLIVER

Predictor CpGs	148
Unit	Day
Target Tissue	liver
Target Population	Mouse
Method	Elastic Net regression
Reference	Wang T, Genome biology (2017).
Alias	

29.1 Usage

```
usage: epical mmLiver [-h] [-o out_prefix] [-g {mm10,mm39}] [-p PERCENT]
                    [-d DELIMITER] [-f {pdf,png}] [-m meta_file] [-l log_file]
                    [--impute {-1,0,1,2,3,4,5,6,7,8,9,10}] [-r ref_file]
                    [--debug] [--overwrite]
                    Input_file
```

positional arguments:

Input_file The input tabular structure file containing DNA methylation data. This file must have a header row, which contains the names or labels for samples. The first column of this file should contain CpG IDs. The remaining cells in the file should contain DNA methylation beta values, represented as floating-point numbers between 0 and 1. Use a TAB, comma, or any other delimiter to separate the columns. Use 'NaN' or 'NA' to represent missing values. This file can be a regular text file or compressed file (".gz", ".Z", ".z", ".bz", ".bz2", ".bzip2").

options:

-h, --help show this help message and exit

-o out_prefix, --output out_prefix
 The PREFIX of output files. If no PREFIX is provided, the default prefix "clock_name_out" is used. The generated output files include:
 "<PREFIX>.DNAm_age.tsv": This file contains the predicted DNAm age. "<PREFIX>.used_CpGs.tsv": This

(continues on next page)

(continued from previous page)

file lists the CpGs that were used to calculate the DNAm age. "<PREFIX>.missed_CpGs.txt": This file provides a list of clock CpGs that were missed or excluded from the input file. "<PREFIX>.coef.tsv": This file contains a list of clock CpGs along with their coefficients. The last column indicates whether the CpG is included in the calculation.

"<PREFIX>.plots.R": This file is an R script used to generate visualization plots.

"<PREFIX>.coef_plot.pdf": This file is the coefficient plot in either PDF or PNG format.

-g {mm10,mm39}, --genome {mm10,mm39}
The reference genome for Mouse (*Mus musculus*) used for RRBS or WGBS reads alignment. Must be 'mm10' or 'mm39'. default is 'mm10'.

-p PERCENT, --percent PERCENT
The maximum allowable percentage of missing CpGs. Set to 0.2 (20%) by default, which means that if more than 20% of the clock CpGs are missing, the estimation of DNAm age cannot be performed.

-d DELIMITER, --delimiter DELIMITER
Separator (usually TAB or comma) used in the input file. If the separator is not provided, the program will automatically detect the separator.

-f {pdf,png}, --format {pdf,png}
The output plot format must be either 'pdf' or 'png'. The default format is 'pdf'."

-m meta_file, --metadata meta_file
This file contains the meta information for each sample. This file must have a header row, which contains the names or labels for variables (such as "Sex", "Age"). If the header row includes an "Age" column, a scatter plot will be generated to display the correlation between chronological age and predicted DNAm age. The first column of this file should contain sample IDs. The default value is None, indicating that no meta information file is provided.

-l log_file, --log log_file
This file is used to save the log information. By default, if no file is specified (None), the log information will be printed to the screen.

--impute {-1,0,1,2,3,4,5,6,7,8,9,10}
The imputation method code must be one of the 12 digits including (-1, 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10). The interpretations are: -1: Remove CpGs with any missing values. 0: Fill all missing values with '0.0'. 1: Fill all missing values with '1.0'. 2: Fill the missing values with ****column mean**** 3: Fill the missing values with ****column median**** 4: Fill the missing values with ****column min**** 5: Fill the missing values with ****column max**** 6: Fill the missing values with ****row mean**** 7: Fill the missing values with

(continues on next page)

(continued from previous page)

```

**row median** 8: Fill the missing values with **row
min** 9: Fill the missing values with **row max** 10:
Fill the missing values with **external reference** If
10 is specified, an external reference file must be
provided.
-r ref_file, --ref ref_file
    The external reference file contains two columns,
    separated by either tabs or commas. The first column
    represents the probe ID, while the second column
    contains the corresponding beta values.
--debug
    If set, print detailed information for debugging.
--overwrite
    If set, over-write existing output files.

```

29.2 Example

```
$ epical mmLiver Test7_mouse_N81_RRBS_beta.tsv -g mm10 -o mmLiver_out
```

```

2024-04-12 12:11:47 [INFO] The prefix of output files is set to "mmLiver_out".
2024-04-12 12:11:47 [INFO] Loading mmLiver clock data ...
2024-04-12 12:11:47 [INFO] Clock's name: "liver_mm10"
2024-04-12 12:11:47 [INFO] Clock was trained from: "Liver"
2024-04-12 12:11:47 [INFO] Clock's unit: "day"
2024-04-12 12:11:47 [INFO] Number of CpGs used: 148
...

```


MMBLOOD

Predictor CpGs	90
Unit	Day
Target Tissue	Blood
Target Population	Mouse
Method	Elastic Net regression
Reference	Petkovich DA, Cell Metab. (2017).
Alias	

30.1 Usage

```
usage: epical mmBlood [-h] [-o out_prefix] [-g {mm10,mm39}] [-p PERCENT]
                    [-d DELIMITER] [-f {pdf,png}] [-m meta_file] [-l log_file]
                    [--impute {-1,0,1,2,3,4,5,6,7,8,9,10}] [-r ref_file]
                    [--debug] [--overwrite]
                    Input_file
```

positional arguments:

Input_file The input tabular structure file containing DNA methylation data. This file must have a header row, which contains the names or labels for samples. The first column of this file should contain CpG IDs. The remaining cells in the file should contain DNA methylation beta values, represented as floating-point numbers between 0 and 1. Use a TAB, comma, or any other delimiter to separate the columns. Use 'NaN' or 'NA' to represent missing values. This file can be a regular text file or compressed file (".gz", ".Z", ".z", ".bz", ".bz2", ".bzip2").

options:

-h, --help show this help message and exit

-o out_prefix, --output out_prefix The PREFIX of output files. If no PREFIX is provided, the default prefix "clock_name_out" is used. The generated output files include:
 "<PREFIX>.DNAm_age.tsv": This file contains the predicted DNAm age. "<PREFIX>.used_CpGs.tsv": This

(continues on next page)

(continued from previous page)

file lists the CpGs that were used to calculate the DNAm age. "<PREFIX>.missed_CpGs.txt": This file provides a list of clock CpGs that were missed or excluded from the input file. "<PREFIX>.coef.tsv": This file contains a list of clock CpGs along with their coefficients. The last column indicates whether the CpG is included in the calculation.

"<PREFIX>.plots.R": This file is an R script used to generate visualization plots.

"<PREFIX>.coef_plot.pdf": This file is the coefficient plot in either PDF or PNG format.

-g {mm10,mm39}, --genome {mm10,mm39}
The reference genome for Mouse (*Mus musculus*) used for RRBS or WGBS reads alignment. Must be 'mm10' or 'mm39'. default is 'mm10'.

-p PERCENT, --percent PERCENT
The maximum allowable percentage of missing CpGs. Set to 0.2 (20%) by default, which means that if more than 20% of the clock CpGs are missing, the estimation of DNAm age cannot be performed.

-d DELIMITER, --delimiter DELIMITER
Separator (usually TAB or comma) used in the input file. If the separator is not provided, the program will automatically detect the separator.

-f {pdf,png}, --format {pdf,png}
The output plot format must be either 'pdf' or 'png'. The default format is 'pdf'.

-m meta_file, --metadata meta_file
This file contains the meta information for each sample. This file must have a header row, which contains the names or labels for variables (such as "Sex", "Age"). If the header row includes an "Age" column, a scatter plot will be generated to display the correlation between chronological age and predicted DNAm age. The first column of this file should contain sample IDs. The default value is None, indicating that no meta information file is provided.

-l log_file, --log log_file
This file is used to save the log information. By default, if no file is specified (None), the log information will be printed to the screen.

--impute {-1,0,1,2,3,4,5,6,7,8,9,10}
The imputation method code must be one of the 12 digits including (-1, 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10). The interpretations are: -1: Remove CpGs with any missing values. 0: Fill all missing values with '0.0'. 1: Fill all missing values with '1.0'. 2: Fill the missing values with ****column mean**** 3: Fill the missing values with ****column median**** 4: Fill the missing values with ****column min**** 5: Fill the missing values with ****column max**** 6: Fill the missing values with ****row mean**** 7: Fill the missing values with

(continues on next page)

(continued from previous page)

```

**row median** 8: Fill the missing values with **row
min** 9: Fill the missing values with **row max** 10:
Fill the missing values with **external reference** If
10 is specified, an external reference file must be
provided.
-r ref_file, --ref ref_file
    The external reference file contains two columns,
    separated by either tabs or commas. The first column
    represents the probe ID, while the second column
    contains the corresponding beta values.
--debug
    If set, print detailed information for debugging.
--overwrite
    If set, over-write existing output files.

```

30.2 Example

```
$ epical mmBlood Test7_mouse_N81_RRBS_beta.tsv -g mm10 -o mmBlood_out
```

```

2024-04-12 12:15:32 [INFO] The prefix of output files is set to "mmBlood_out".
2024-04-12 12:15:32 [INFO] Loading mmBlood clock data ...
2024-04-12 12:15:32 [INFO] Clock's name: "blood_mm10"
2024-04-12 12:15:32 [INFO] Clock was trained from: "Blood"
2024-04-12 12:15:32 [INFO] Clock's unit: "day"
2024-04-12 12:15:32 [INFO] Number of CpGs used: 90
...

```


LICENSE

Copyright (c) 2024 The Python Packaging Authority

Permission is hereby granted, free of charge, to any person obtaining a copy of this software and associated documentation files (the “Software”), to deal in the Software without restriction, including without limitation the rights to use, copy, modify, merge, publish, distribute, sublicense, and/or sell copies of the Software, and to permit persons to whom the Software is furnished to do so, subject to the following conditions:

The above copyright notice and this permission notice shall be included in all copies or substantial portions of the Software.

THE SOFTWARE IS PROVIDED “AS IS”, WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO THE WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT. IN NO EVENT SHALL THE AUTHORS OR COPYRIGHT HOLDERS BE LIABLE FOR ANY CLAIM, DAMAGES OR OTHER LIABILITY, WHETHER IN AN ACTION OF CONTRACT, TORT OR OTHERWISE, ARISING FROM, OUT OF OR IN CONNECTION WITH THE SOFTWARE OR THE USE OR OTHER DEALINGS IN THE SOFTWARE.

REFERENCE

Epical: an epigenetic age calculator. (In preparation).

Symbols

1. EPM.train_selected_CpGs.tsv, [60](#)
2. EPM.train_EPM_age.tsv, [60](#)
3. EPM.train_EPM_age.pdf, [60](#)
4. EPM.test_selected_CpGs.tsv, [60](#)
5. EPM.test_EPM_age.tsv, [60](#)
6. EPM.test_EPM_age.pdf, [60](#)