**Supplementary Information**

**Prediction of biological age and evaluation of genome-wide dynamic methylomic changes throughout human aging**

Mahmoud Amiri Roudbar, Seyedeh Fatemeh Mousavi, Siavash Salek Ardestani, Fernando Brito Lopes, Mehdi Momen, Daniel Gianola, Hasan Khatib.



**Figure S1.** A histogram of the age distribution for all individuals.



**Figure S2.** Scatter plot of methylomic predicted age for all individuals based on the Bayes Ridge Regression (BRR) and reproducing kernel Hilbert space (RKHS) approaches using a subset of the age-associated methylation sites. Correlation between predicted age using BRR and RKHS is close to one with a very small root mean square error (RMSE).

****

**Figure S3.** Plots of methylation beta values for markers with (a) and without (b) heteroscedasticity. Plots show changes in methylation residuals with age, defined as the difference between an individual’s methylation beta value and expected methylation beta value (blue line).

**Table S1.** Details of each dataset.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Availability** | **DNA origin** | **Sample size (Prob. Female)** | **Median age (range)** | **Mean age (SD)** | **Citation** |
| **GSE55763** | Peripheral blood | 2675 (0.32) | 50.1 (23.7, 75.0) | 51 (10.09) | Lehne et al. 2015 |
| **GSE40279** | Whole blood | 656 (0.52) | 65 (19, 101) | 64 (14.74) | Hannum et al. 2013 |
| **GSE56105** | Whole blood | 614 (0.49) | 14 (10, 75) | 21.3 (14.11) | McRae et al. 2014 |
| **GSE50660** | Peripheral blood | 464 (0.3) | 56 (38, 67) | 55.4 (6.66) | Tsaprouni et al. 2014 |

**Table S2.** The estimated parameters from the linear model for age with sex and dataset as fixed effects and B cell type as a covariate.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Estimate** | **Std. Error** | **t value** | **Pr(>|t|)** |
| **Intercept** | 64.352 | 0.574 | 112.026 | 0.00E+00 |
| **Male** | -0.89 | 0.356 | -2.5 | 1.24E-02 |
| **Dataset 'GSE50660'** | -8.459 | 0.687 | -12.322 | 2.52E-34 |
| **Dataset 'GSE55763'** | -12.882 | 0.495 | -26.046 | 3.21E-139 |
| **Dataset 'GSE56105'** | -42.698 | 0.632 | -67.587 | 0.00E+00 |
| **B cell** | 1.722 | 5.02 | 0.343 | 7.32E-01 |

**Table S3.** The estimated parameters from the linear model for age with sex and dataset as fixed effects and CD4T cell type as a covariate.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Estimate** | **Std. Error** | **t value** | **Pr(>|t|)** |
| **Intercept** | 64.443 | 0.62 | 103.983 | 0.00E+00 |
| **Male** | -0.886 | 0.356 | -2.49 | 1.28E-02 |
| **Dataset 'GSE50660'** | -8.454 | 0.686 | -12.317 | 2.67E-34 |
| **Dataset 'GSE55763'** | -12.881 | 0.495 | -26.043 | 3.47E-139 |
| **Dataset 'GSE56105'** | -42.711 | 0.636 | -67.174 | 0.00E+00 |
| **CD4T cell** | 0.154 | 2.913 | 0.053 | 9.58E-01 |

**Table S4.** The estimated parameters from the linear model for age with sex and dataset as fixed effects and CD8T cell type as a covariate.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Estimate** | **Std. Error** | **t value** | **Pr(>|t|)** |
| **Intercept** | 64.342 | 0.551 | 116.871 | 0.00E+00 |
| **Male** | -0.89 | 0.356 | -2.499 | 1.25E-02 |
| **Dataset 'GSE50660'** | -8.46 | 0.686 | -12.323 | 2.49E-34 |
| **Dataset 'GSE55763'** | -12.872 | 0.495 | -26.005 | 8.09E-139 |
| **Dataset 'GSE56105'** | -42.677 | 0.635 | -67.195 | 0.00E+00 |
| **CD8T cell** | 1.639 | 3.803 | 0.431 | 6.67E-01 |

**Table S5.** The estimated parameters from the linear model for age with sex and dataset as fixed effects and granulocyte cell type as a covariate.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Estimate** | **Std. Error** | **t value** | **Pr(>|t|)** |
| **Intercept** | 64.353 | 1.157 | 55.612 | 0.00E+00 |
| **Male** | -0.884 | 0.356 | -2.481 | 1.31E-02 |
| **Dataset 'GSE50660'** | -8.454 | 0.686 | -12.316 | 2.71E-34 |
| **Dataset 'GSE55763'** | -12.883 | 0.495 | -26.012 | 7E-139 |
| **Dataset 'GSE56105'** | -42.71 | 0.632 | -67.554 | 0.00E+00 |
| **Granulocyte** | 0.196 | 1.853 | 0.106 | 9.16E-01 |

**Table S6.** The estimated parameters from the linear model for age with sex and dataset as fixed effects and monocyte cell type as a covariate.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Estimate** | **Std. Error** | **t value** | **Pr(>|t|)** |
| **Intercept** | 65.106 | 0.642 | 101.369 | 0.00E+00 |
| **Male** | -0.884 | 0.356 | -2.486 | 1.30E-02 |
| **Dataset 'GSE50660'** | -8.444 | 0.686 | -12.305 | 3.10E-34 |
| **Dataset 'GSE55763'** | -12.864 | 0.495 | -26.01 | 7.34E-139 |
| **Dataset 'GSE56105'** | -42.672 | 0.632 | -67.564 | 0.00E+00 |
| **Monocyte** | -10.093 | 6.864 | -1.47 | 1.42E-01 |

**Table S7.** The estimated parameters from the linear model for age with sex and dataset as fixed effects and NK cell type as a covariate.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Estimate** | **Std. Error** | **t value** | **Pr(>|t|)** |
| **Intercept** | 64.499 | 0.553 | 116.697 | 0.00E+00 |
| **Male** | -0.884 | 0.356 | -2.484 | 1.30E-02 |
| **Dataset 'GSE50660'** | -8.456 | 0.687 | -12.317 | 2.68E-34 |
| **Dataset 'GSE55763'** | -12.886 | 0.497 | -25.948 | 2.95E-138 |
| **Dataset 'GSE56105'** | -42.718 | 0.638 | -66.944 | 0.00E+00 |
| **NK cell** | -0.395 | 3.325 | -0.119 | 9.06E-01 |

**Table S8.** The estimated parameters from the linear model for age with sex and dataset as fixed effects and B cell type as a covariate.

|  |  |  |  |
| --- | --- | --- | --- |
| **Method** | **No. of the CpG sites in the original model** | **No. of the missed CpG sites** | **Name of the missed CpG sites** |
| Horvath multi-tissue | 353 | 2 | cg05755779; cg25683012 |
| Horvath skin | 391 | 9 | cg18769120; cg23091758; cg26311454; cg13027206; cg25005357; cg11620135; cg21944491; cg21911021; cg01892695 |
| Levine | 513 | 4 | cg18392482; cg01128603; cg20674577; cg08212685 |

**Table S9.** Results of the Breusch-Pagan test for heteroscedasticity of aDMPs from un-adjusted beta values.

|  |  |  |
| --- | --- | --- |
|  | **Number of markers with heteroscedasticity (% of increased with aging)** | **Number of markers without heteroscedasticity** |
| **All markers** | 121023 (97.29) | 2907 |
| **CpG Island** | 35122 (97.91) | 753 |
| **CpG Shore** | 31132 (96.83) | 671 |
| **CpG Shelf** | 11136 (97.41) | 297 |
| **CpG Open Sea** | 43633 (97.1) | 1186 |
| **Enhancer** | 35279 (97.65) | 832 |
| **DHS** | 21347 (97.41) | 491 |
| **CDMR; DMR; RDMR1** | 2840 (98.8); 9835 (99.08); 5204 (97.67) | 48; 186; 91 |
| **Regulatory associated** | 37292 (95.4) | 1110 |

 1DMR, Differentially Methylated Region; CDMR, Cancer-specific Differentially Methylated Region; RDMR, Reprogramming-specific Differentially Methylated Region.

**Table S10.** Age-associated methylation sites properties from the adjusted and un-adjusted beta values.

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Total number of markers** | **Number of associated markers in adjusted (% of positively associated)** | **Number of associated markers in un-adjusted (% of positively associated)** |
| **All associated markers** | 423394 | 107806 (47.25) | 123930 (41.11) |
| **CpG Island** | 150254 | 35510 (82.86) | 35875 (82.76) |
| **CpG Shore** | 112067 | 27438 (45.6) | 31803 (38.84) |
| **CpG Shelf** | 47144 | 8873 (16.7) | 11433 (14.2) |
| **CpG Open Sea** | 176047 | 35985 (20.9) | 44819 (16.24) |
| **Enhancer** | 102559 | 30174 (41.19) | 36111 (34.12) |
| **DHS** | 59916 | 19919 (69.6) | 21838 (62.89) |
| **CDMR; DMR; RDMR1** | 6659; 18447; 12231 | 2507 (66.57); 9767 (94.23); 4770 (60.86) | 2888 (62.53); 10021 (93.85); 5295 (55.6) |
| **Regulatory associated** | 170092 | 35542 (54.66) | 38402 (51.41) |
| **5' UTR** | 65535 | 14202 (56.63) | 15640 (51.53) |
| **1st exon** | 39372 | 8732 (69.42) | 8979 (67.6) |

1DMR, Differentially Methylated Region; CDMR, Cancer-specific Differentially Methylated Region; RDMR, Reprogramming-specific Differentially Methylated Region.

**Table S11.** The estimated parameters of the linear model for the average of un-adjusted methylation from the associated markers located at the CpG islands with sex, dataset, and age as fixed effects.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Estimate** | **Std. Error** | **t value** | **Pr(>|t|)** |
| **Intercept** | 1.77E-01 | 9.37E-04 | 188.89 | 0.00E+00 |
| **Age** | 3.46E-04 | 1.31E-05 | 26.48 | 1.59E-143 |
| **Dataset 'GSE50660'** | 6.25E-03 | 6.06E-04 | 10.32 | 1.10E-24 |
| **Dataset 'GSE55763'** | 9.91E-03 | 4.61E-04 | 21.5 | 1.48E-97 |
| **Dataset 'GSE56105'** | 6.96E-03 | 7.82E-04 | 8.91 | 7.58E-19 |
| **Male** | 9.50E-04 | 3.09E-04 | 3.07 | 2.12E-03 |

**Table S12.** The estimated parameters of the linear model for the average of un-adjusted methylation from the associated markers located at the CpG shores with sex, dataset, and age as fixed effects.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Estimate** | **Std. Error** | **t value** | **Pr(>|t|)** |
| **Intercept** | 4.29E-01 | 6.35E-04 | 674.99 | 0.00E+00 |
| **Age** | -1.23E-04 | 8.87E-06 | -13.92 | 4.02E-43 |
| **Dataset 'GSE50660'** | 3.69E-03 | 4.11E-04 | 8.98 | 3.83E-19 |
| **Dataset 'GSE55763'** | 4.04E-03 | 3.13E-04 | 12.93 | 1.42E-37 |
| **Dataset 'GSE56105'** | 2.48E-03 | 5.30E-04 | 4.69 | 2.88E-06 |
| **Male** | -6.21E-04 | 2.09E-04 | -2.97 | 3.03E-03 |

**Table S13.** The estimated parameters of the linear model for the average of un-adjusted methylation from the associated markers located at the CpG shelves with sex, dataset, and age as fixed effects.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Estimate** | **Std. Error** | **t value** | **Pr(>|t|)** |
| **Intercept** | 6.75E-01 | 7.27E-04 | 928.93 | 0.00E+00 |
| **Age** | -3.00E-04 | 1.01E-05 | -29.61 | 7.41E-176 |
| **Dataset 'GSE50660'** | -7.71E-03 | 4.70E-04 | -16.41 | 8.40E-59 |
| **Dataset 'GSE55763'** | -7.21E-03 | 3.58E-04 | -20.15 | 1.99E-86 |
| **Dataset 'GSE56105'** | -1.74E-03 | 6.07E-04 | -2.86 | 4.27E-03 |
| **Male** | -9.76E-04 | 2.40E-04 | -4.07 | 4.78E-05 |

**Table S14.** The estimated parameters of the linear model for the average of un-adjusted methylation from the associated markers located at the CpG open seas with sex, dataset, and age as fixed effects.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Estimate** | **Std. Error** | **t value** | **Pr(>|t|)** |
| **Intercept** | 6.57E-01 | 7.40E-04 | 887.07 | 0.00E+00 |
| **Age** | -2.93E-04 | 1.03E-05 | -28.4 | 4.65E-163 |
| **Dataset 'GSE50660'** | -6.22E-03 | 4.78E-04 | -13 | 6.24E-38 |
| **Dataset 'GSE55763'** | -9.15E-03 | 3.64E-04 | -25.14 | 1.99E-130 |
| **Dataset 'GSE56105'** | -1.58E-03 | 6.18E-04 | -2.56 | 1.04E-02 |
| **Male** | -5.84E-04 | 2.44E-04 | -2.39 | 1.68E-02 |

**Table S15.** Diseases associated with the set of genes that were determined to have age-associated methylation alterations using genetic association database (GAD).

|  |  |  |  |
| --- | --- | --- | --- |
| **GAD term** | **Count** | **%** | **FDR** |
| Tobacco Use Disorder | 2177 | 20.62 | 8.39E-138 |
| Hip | 241 | 2.28 | 2.10E-05 |
| Body Mass Index | 308 | 2.92 | 6.36E-05 |
| Erythrocyte Count | 136 | 1.29 | 2.86E-04 |
| Cholesterol | 275 | 2.60 | 3.01E-04 |
| Body Height | 356 | 3.37 | 6.39E-04 |
| Waist Circumference | 170 | 1.61 | 9.06E-04 |
| Heart Rate | 178 | 1.69 | 1.12E-03 |
| Heart Failure | 248 | 2.35 | 1.26E-03 |
| Type 2 Diabetes | 1244 | 11.78 | 2.26E-03 |
| Hemoglobins | 173 | 1.64 | 4.08E-03 |
| Triglycerides | 275 | 2.60 | 5.41E-03 |
| Arteries | 112 | 1.06 | 6.31E-03 |
| Celiac Disease | 191 | 1.81 | 8.97E-03 |
| Coronary Artery Disease | 315 | 2.98 | 1.05E-02 |
| Body Weight | 258 | 2.44 | 1.13E-02 |
| Cell Adhesion Molecules | 53 | 0.50 | 1.19E-02 |
| Stroke | 332 | 3.14 | 1.62E-02 |
| Alcoholism | 312 | 2.96 | 2.88E-02 |
| Blood Pressure | 285 | 2.70 | 3.51E-02 |
| Iron | 169 | 1.60 | 4.27E-02 |
| Parkinson Disease | 142 | 1.34 | 4.55E-02 |

**Table S16.** KEGG Pathway associated with our set of genes that were determined to have age-associated methylation alterations.

|  |  |  |  |
| --- | --- | --- | --- |
| **KEGG pathway term** | **Count** | **%** | **FDR** |
| hsa04510:Focal adhesion | 147 | 1.39 | 3.35E-08 |
| hsa04810:Regulation of actin cytoskeleton | 142 | 1.34 | 2.53E-05 |
| hsa04724:Glutamatergic synapse | 83 | 0.79 | 2.99E-04 |
| hsa04152:AMPK signaling pathway | 88 | 0.83 | 4.38E-04 |
| hsa04728:Dopaminergic synapse | 90 | 0.85 | 1.09 E-03 |
| hsa04014:Ras signaling pathway | 146 | 1.38 | 1.14 E-03 |
| hsa04071:Sphingolipid signaling pathway | 85 | 0.80 | 1.42 E-03 |
| hsa04151:PI3K-Akt signaling pathway | 211 | 2.00 | 1.54 E-03 |
| hsa04520:Adherens junction | 55 | 0.52 | 2.09E-03 |
| hsa05100:Bacterial invasion of epithelial cells | 59 | 0.56 | 3.10E-03 |
| hsa04022:cGMP-PKG signaling pathway | 106 | 1.00 | 3.54E-03 |
| hsa05200:Pathways in cancer | 235 | 2.23 | 4.10E-03 |
| hsa04012:ErbB signaling pathway | 64 | 0.61 | 5.25E-03 |
| hsa05231:Choline metabolism in cancer | 72 | 0.68 | 7.59E-03 |
| hsa04010:MAPK signaling pathway | 158 | 1.50 | 7.60E-03 |
| hsa04919:Thyroid hormone signaling pathway | 80 | 0.76 | 9.36E-03 |
| hsa04261:Adrenergic signaling in cardiomyocytes | 93 | 0.88 | 1.20E-02 |
| hsa04015:Rap1 signaling pathway | 133 | 1.26 | 1.68E-02 |
| hsa04911:Insulin secretion | 61 | 0.58 | 3.28E-02 |
| hsa04512:ECM-receptor interaction | 62 | 0.59 | 3.91E-02 |

 **Table S17.** Estimates of regression coefficient and their standard error (in parentheses) and AIC from models without or with sex effect (Model 1 and 2, respectively) and model with interaction between age and sex (Model 3) for AMAR in reproducing kernel Hilbert space (RKHS) and Bayesian Ridge Regression (BRR) approaches; best model is displayed in bold according to AIC estimates*.*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Approach** | **Model** | **Intercept (SE)** | **Reg. Coef. of Age (SE)** | **Reg. Coef. of Male (SE)** | **Reg. Coef. of Male:age (SE)** | **AIC** |
| **RKHS** | Model 1 | 1.01e+00 (6.44e-04) | -1.14e-04 (1.24e-05) | - | - | -25369.55 |
| Model 2 | 1.01e+00 (6.96e-04) | -1.14e-04 (1.24e-05) | -4.19e-04 (4.24e-04) | - | -25368.53 |
| **Model 3** | **1.01e+00 (9.77e-04)** | **-1.60e-04 (1.87e-05)** | **-4.46e-03 (1.30e-03)** | **8.21e-05 (2.50e-05)** | **-25377.35** |
| **BRR** | Model 1 | 1.01e+00 (6.46e-04) | -9.56e-05 (1.24e-05) | - | - | -25342.9 |
| Model 2 | 1.01e+00 (6.98e-04) | -9.55e-05 (1.24e-05) | -5.02e-04 (4.26e-04) | - | -25342.29 |
| **Model 3** | **1.01e+00 (9.79e-04)** | **-1.44e-04 (1.87e-05)** | **-4.73e-03 (1.30e-03)** | **8.60e-05 (2.50e-05)** | **-25352.09** |